Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/EP05/050900

International filing date: 01 March 2005 (01.03.2005)

Document type: Certified copy of priority document

Document details: Country/Office: EP
Number: 04405124.1
Filing date: 02 March 2004 (02.03.2004)

Date of receipt at the International Bureau: 19 April 2005 (19.04.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

Europäisches Patentamt | European Patent Office | Office européen des brevets

01. 03. 2005

Bescheinigung | Certificate | Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

04405124.1

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Anmeldung Nr:
Application no.: 04405124.1
Demande no:

Anmeldetag:
Date of filing: 02.03.04
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Ecole Polytechnique Fédérale de Lausanne
CM-Ecublens
1015 Lausanne
SUISSE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Specific substrates for O6-alkylguanine-DNA alkyltransferase

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

C07D473/00

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL PL PT RO SE SI SK TR LI

Specific substrates for O^6-Alkylguanine-DNA Alkyltransferase

Field of the Invention

The present invention relates to methods of transferring a label from substrates to O^6-alkylguanine-DNA alkyltransferases (AGT) and O^6-alkylguanine-DNA alkyltransferase fusion proteins, and to novel specific substrates suitable in such methods.

Background of the invention

The mutagenic and carcinogenic effects of electrophiles such as N-methyl-N-nitrosourea are mainly due to the O^6-alkylation of guanine in DNA. To protect themselves against DNA-alkylation, mammals and bacteria possess a protein, O^6-alkylguanine-DNA alkyltransferase (AGT) which repairs these lesions. AGT transfers the alkyl group from the position O^6 of alkylated guanine and guanine derivatives to the mercapto group of one of its own cysteines, resulting in an irreversibly alkylated AGT. The underlying mechanism is a nucleophilic reaction of the S_N2 type which explains why not only methyl groups, but also benzylic groups are easily transferred. As overexpression of AGT in tumour cells is the main reason for resistance to alkylating drugs such as procarbazine, dacarbazine, temozolomide and bis-2-chloroethyl-N-nitrosourea, inhibitors of AGT have been proposed for use as sensitisers in chemotherapy (Pegg *et al.*, Prog Nucleic Acid Res Mol Biol 51:167-223, 1995). US 5,691,307 describes O^6-benzylguanines carrying various substituents in the benzyl group, and their use for depleting AGT levels in tumor cells and thereby increasing responsiveness to alkylating anti-tumor drugs. Likewise, WO 97/20843 discloses further AGT depleting compounds representing O^6-benzyl- and O^6-heteroarylmethyl-pyrimidine derivatives.

DE 199 03 895 discloses an assay for measuring levels of AGT which relies on the reaction between biotinylated O^6-alkylguanine derivatives and AGT which leads to biotinylation of the AGT. This in turn allows the separation of the AGT on a streptavidin coated plate and its detection, e.g. in an ELISA assay. The assay is suggested for monitoring the level of AGT in tumour tissue and for use in screening for AGT inhibitors.

WO 01/85221 proposes the use of radiolabelled fluoro- or iodo-substituted O^6-benzyl-guanines for detection of AGT and monitoring the level of AGT.

Damoiseaux *et al.,* ChemBiochem. 4:285-287, 2001, disclose modified O^6-alkylated guanine derivatives incorporated into oligodeoxyribonucleotides for use as chemical probes for labeling AGT, again to facilitate detecting the levels of this enzyme in cancer cells to aid in research and in chemotherapy.

WO 02/083937 discloses a method for detecting and/or manipulating a protein of interest wherein the protein is fused to AGT and the AGT fusion protein contacted with an AGT substrate carrying a label, and the AGT fusion protein detected and optionally further manipulated using the label. Several AGT fusion proteins to be used, general structural principles of the AGT substrate and a broad variety of labels and methods to detect the label useful in the method are described.

PCT/EP03/10859 describes particular AGT fusion proteins to be used in the mentioned method for detecting and/or manipulating a protein of interest, labelled fusion proteins obtainable by this method, and the method using the particular AGT fusion proteins.

PCT/EP03/10889 discloses additional AGT substrates carrying a label particularly suitable in the mentioned method for detecting and/or manipulating a protein of interest, and the application of such particularly labelled substrates. This patent application also describes methods of manufacture of these additional AGT substrates.

Summary of the invention

The invention relates to substrates for O^6-alkylguanine-DNA alkyltransferases (AGT) of formula (1)

$$R_1\text{–}A\text{–}X\text{–}CH_2\text{–}R_3\text{–}R_4\text{–}L_1 \qquad (1)$$

wherein A is a group recognized by AGT as a substrate;
X is oxygen or sulfur;
R_1 is a group $-R_2-L_2$ or a group R_5;

R_2 and R_4 are, independently of each other, a linker;
R_3 is an aromatic or a heteroaromatic group, or an optionally substituted unsaturated alkyl, cycloalkyl or heterocyclyl group with the double bond connected to CH_2;
R_5 is arylmethyl or heteroarylmethyl or an optionally substituted cycloalkyl, cycloalkenyl or heterocyclyl group;
L_1 is a label, a plurality of same or different labels, a bond connecting R_4 to A forming a cyclic substrate, or a further group $-R_3-CH_2-X-A-R_1$; and
L_2 is a label or a plurality of same or different labels.

The invention further relates to methods of transferring a label from these substrates to O^6-alkylguanine-DNA alkyltransferases (AGT) and AGT fusion proteins.

Brief description of the Figures

Figure 1: Schematic representation of the reaction of fusion protein comprising a protein of interest (P) and AGT with substrates of formula $R_1-A-X-CH_2-R_3-R_4-L_1$.

Figure 2: Western blot for the AGTM-CPTG kinetic measurement: Intensity of the AGTM-BGBT complex detected with luminescent peroxidase substrate, see Example 8. Aliquots of mutant AGTM-6xHis are incubated with 10fold excess of N^9-cyclopentyl-O^6-(4-bromothiophen-2-yl)-guanine (CPTG) in aqueous buffered solution. After a predetermined time, the reaction is quenched with a biotinylated O^6-benzylguanine (BGBT, substance 3a of Juillerat *et al.*, Chem Biol 10:313-317, 2003), which is supposed to react much faster with the protein, thus capturing unreacted AGTM. Proteins are then denatured with SDS and heat. Samples are subjected to SDS-PAGE and Western blot analysis. The intensity of the corresponding bands is detected by chemiluminescence stain.

Figure 3: Western blot of the reaction of the 2 proteins (AGTM-6xHis at ~ 25 kDa, 0.2 µM, and wild type hAGT-GST fusion protein at ~ 50 kDa, 1.2 µM) with the two substrates N^9-cyclopentyl-O^6-(4-bromothiophen-2-yl)-guanine (CPTG, 0, 5 and 10 µM) and biotinylated O^6-benzylguanine (BGBT, 5 µM), see Example 9. Detection of the AGTM-BGBT complex with luminescent peroxidase substrate is as described for Figure 2.

Detailed description of the invention

The particular AGT substrates of the invention are compounds of formula (1)

$$R_1–A–X–CH_2–R_3–R_4–L_1 \qquad (1)$$

wherein A is a group recognized by AGT as a substrate;
X is oxygen or sulfur;
R_1 is a group $–R_2–L_2$ or a group R_5;
R_2 and R_4 are, independently of each other, a linker;
R_3 is an aromatic or a heteroaromatic group, or an optionally substituted unsaturated alkyl, cycloalkyl or heterocyclyl group with the double bond connected to CH_2;
R_5 is arylmethyl or heteroarylmethyl or an optionally substituted cycloalkyl, cycloalkenyl or heterocyclyl group,
L_1 is a label, a plurality of same or different labels, a bond connecting R_4 to A forming a cyclic substrate, or a further group $–R_3–CH_2–X–A–R_1$; and
L_2 is a label or a plurality of same or different labels.

In a group $R_1–A$, the residue A is preferably a heteroaromatic group containing 1 to 5 nitrogen atoms, recognized by AGT as a substrate.

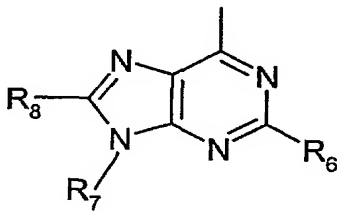
A heteroaromatic group A is mono- or bicyclic and has 5 to 12, preferably 6 or 9 or 10 ring atoms; which in addition to carrying a substituent R_1 may be unsubstituted or substituted by one or more, especially one, two or three further substituents selected from the group consisting of lower alkyl, such as methyl, lower alkoxy, such as methoxy or ethoxy, hydroxy, oxo, amino, lower alkylamino, di-lower alkylamino, acylamino, halogen, such as chlorine or bromine, halogenated lower alkyl, such as trifluoromethyl, carboxy, lower alkoxycarbonyl, carbamoyl, lower alkylcarbamoyl, or lower alkylcarbonyl.

Lower alkyl is preferably alkyl with 1 to 7, preferably from 1 to 4 C atoms, and is linear or branched; preferably, lower alkyl is butyl, such as n-butyl, sec-butyl, isobutyl, tert-butyl, propyl, such as n-propyl or isopropyl, ethyl or methyl. Most preferably, lower alkyl is methyl.

In lower alkoxy, the lower alkyl group is as defined hereinbefore. Lower alkoxy denotes preferably n-butoxy, tert-butoxy, iso-propoxy, ethoxy, or methoxy, in particular methoxy.

Preferably the mono- or bicyclic heteroaromatic group A is selected from 2H-pyrrolyl, pyrrolyl, imidazolyl, benzimidazolyl, pyrazolyl, indazolyl, purinyl, 8-azapurinyl, 7-deazapurinyl, 8-aza-7-deazapurinyl, pyridyl, pyrazinyl, pyrimidinyl, pyridazinyl, triazinyl, 4H-quinolizinyl, isoquinolyl, quinolyl, phthalazinyl, naphthyridinyl, quinoxalyl, quinazolinyl, quinolinyl, pteridinyl, indolizinyl, 3H-indolyl, indolyl, isoindolyl, triazolyl, tetrazolyl, or benzo[d]pyrazolyl. More preferably the mono- or bicyclic heteroaromatic group A is selected from the group consisting of purinyl, 8-azapurinyl, 7-deazapurinyl, 8-aza-7-deazapurinyl, pyridyl, pyrazinyl, pyrimidinyl, pyridazinyl and triazinyl.

For example the group R_1–A may be a purine radical of formula (2)



(2)

wherein R_6 is hydrogen, hydroxy or unsubstituted or substituted amino; and one of R_7 and R_8 is R_1 and the other one is hydrogen.

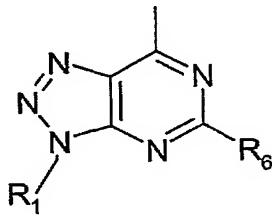
If R_6 is hydroxy, the purine radical is predominantly present in its tautomeric form wherein a nitrogen adjacent to the carbon atom bearing R_6 carries a hydrogen atom, the double bond between this nitrogen atom and the carbon atom bearing R_6 is a single bond, and R_6 is double bonded oxygen, respectively.

A substituted amino group R_6 is lower alkylamino of 1 to 4 carbon atoms or acylamino, wherein the acyl group is lower alkylcarbonyl with 1 to 5 carbon atoms, e.g. acetyl, propionyl, n- or isopropylcarbonyl, or n-, iso- or tert-butylcarbonyl, or arylcarbonyl, e.g. benzoyl.

If R_6 is unsubstituted or substituted amino and the residue X connected to the bond of the purine radical is oxygen, the residue of formula (2) is a guanine derivative.

Particularly preferred are compounds wherein the group R_1–A is a purine radical of formula (2), R_6 is unsubstituted amino, R_7 is R_1, R_8 is hydrogen, and X is oxygen, i.e. a guanine derivative carrying a further substituent in position N^9.

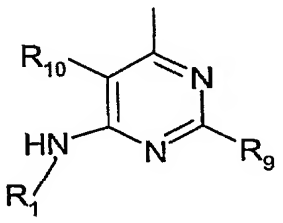
In another preferred embodiment of the invention the group R_1–A is an 8-azapurine radical of formula (3)



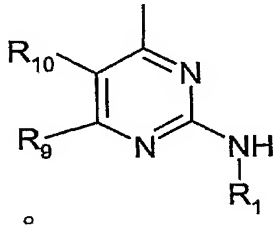
(3)

wherein the substituents R_6 has the meaning as defined for R_6 under formula (2).

In a further preferred embodiment of the invention the group R_1–A is a pyrimidine radical of formula (4a) or (4b)



(4a)



(4b)

wherein R_9 is hydrogen, halogen, lower alkyl with 1 to 4 carbon atoms or amino, preferably amino, and R_{10} is hydrogen, halogen, lower alkyl with 1 to 4 carbon atoms, amino, nitro or nitroso. Halogen R_9 or R_{10} is e.g. fluoro, chloro, bromo or iodo.

X is preferably oxygen.

A linker group R_2 or R_4 is preferably a flexible linker connecting a label L_2 or L_1, respectively, or a plurality of same or different labels L_2 or L_1 to the substrate. Linker units are chosen in the context of the envisioned application, i.e. in the transfer of the substrate to a fusion protein comprising AGT. They also increase the solubility of the substrate in the appropriate solvent. The linkers used are chemically stable under the conditions of the actual application. The linker does not interfere with the reaction with AGT nor with the detection of the label L_1 and/or L_2, but may be constructed such as to be cleaved at some point in time after the reaction of the compound of formula (1) with the fusion protein comprising AGT.

A linker R_2 or R_4 is a straight or branched chain alkylene group with 1 to 300 carbon atoms, wherein optionally

(a) one or more carbon atoms are replaced by oxygen, in particular wherein every third carbon atom is replaced by oxygen, e.g. a poylethyleneoxy group with 1 to 100 ethyleneoxy units;

(b) one or more carbon atoms are replaced by nitrogen carrying a hydrogen atom, and the adjacent carbon atoms are substituted by oxo, representing an amide function –NH–CO–;

(c) one or more carbon atoms are replaced by oxygen, and the adjacent carbon atoms are substituted by oxo, representing an ester function –O–CO–;

(d) the bond between two adjacent carbon atoms is a double or a triple bond, representing a function –CH=CH– or –C≡C–;

(e) one or more carbon atoms are replaced by a phenylene, a saturated or unsaturated cycloalkylene, a saturated or unsaturated bicycloalkylene, a bridging heteroaromatic or a bridging saturated or unsaturated heterocyclyl group;

(f) two adjacent carbon atoms are replaced by a disulfide linkage –S–S–;

or a combination of two or more, especially two or three, alkylene and/or modified alkylene groups as defined under (a) to (f) hereinbefore, optionally containing substituents.

Substituents considered are e.g. lower alkyl, e.g. methyl, lower alkoxy, e.g. methoxy, lower acyloxy, e.g. acetoxy, or halogenyl, e.g. chloro.

Further substituents considered are e.g. those obtained when an α-amino acid, in particular a naturally occurring α-amino acid, is incorporated in the linker R_2 or R_4 wherein carbon atoms are replaced by amide functions–NH–CO– as defined under (b). In such a linker, part of the carbon chain of the alkylene group R_2 or R_4 is replaced by a group $–(NH\text{-}CHR\text{-}CO)_n–$ wherein n is between 1 and 100 and R represents a varying residue of an α-amino acid.

A further substituent is one which leads to a photocleavable linker R_2 or R_4, e.g. an o-nitrophenyl group. In particular this substituent o-nitrophenyl is located at a carbon atom adjacent to an amide bond, e.g. in a group $–NH–CO–CH_2–CH$(o-nitrophenyl)$–NH–CO–$, or as a substituent in a polyethylene glycol chain, e.g. in a group $–O–CH_2–CH$(o-nitrophenyl)$–O–$. Other photocleavable linkers considered are e.g. phenacyl, alkoxybenzoin, benzylthioether and pivaloyl glycol derivatives.

A phenylene group replacing carbon atoms as defined under (e) hereinbefore is e.g. 1,2-, 1,3-, or preferably 1,4-phenylene. In a particular embodiment, the phenylene group is further substituted by a nitro group, and, combined with other replacements as mentioned above under (a), (b), (c), (d), and (f), represents a photocleavable group, and is e.g. 4-nitro-1,3-phenylene, such as in $–CO–NH–CH_2–$4-nitro-1,3-phenylene$–CH(CH_3)–O–CO–$, or 2-methoxy-5-nitro-1,4-phenylene, such as in $–CH_2–O–$2-methoxy-5-nitro-1,4-phenylene$–CH(CH_3)–O–$. Other particular embodiments representing photocleavable linkers are e.g. $–$1,4-phenylene$–CO–CH_2–O–CO–CH_2–$ (a phenacyl group), –1,4-phenylene–CH(OR)–CO–1,4-phenylene– (an alkoxybenzoin), or –3,5-dimethoxy-1,4-phenylene$–CH_2–O–$ (a dimethoxybenzyl moiety). A saturated or unsaturated cycloalkylene group replacing carbon atoms as defined under (e) hereinbefore is derived from cycloalkyl with 3 to 7 carbon atoms, preferably from cyclopentyl or cyclohexyl, and is e.g. 1,2- or 1,3-cyclopentylene, 1,2-, 1,3-, or preferably 1,4-cyclohexylene, or also 1,4-cyclohexylene being unsaturated e.g. in 1- or in 2-position. A saturated or unsaturated bicycloalkylene group replacing carbon atoms as defined under (e) hereinbefore is derived from bicycloalkyl with 7 or 8 carbon atoms, and is e.g. bicyclo[2.2.1] heptylene or bicyclo[2.2.2]octylene, preferably 1,4-bicyclo[2.2.1]-heptylene optionally unsaturated in 2-position or doubly unsaturated in 2- and 5-position, and 1,4-bicyclo[2.2.2]octylene optionally unsaturated in 2-position or doubly unsaturated in 2- and 5-position. A bridging heteroaromatic group replacing carbon atoms as defined

under (e) hereinbefore is e.g. triazolidene, preferably 1,4-triazolidene, or isoxazolidene, preferably 3,5-isoxazolidene. A bridging saturated or unsaturated heterocyclyl group replacing carbon atoms as defined under (e) hereinbefore is e.g. derived from an unsaturated heterocyclyl group as defined under R_3 above, e.g. isoxazolidinene, preferably 3,5-isoxazolidinene, or a fully saturated heterocyclyl group with 3 to 12 atoms, 1 to 3 of which are heteroatoms selected from nitrogen, oxygen and sulfur, e.g. pyrrolidinediyl, piperidinediyl, tetrahydrofuranediyl, dioxanediyl, morpholinediyl or tetrahydrothiophenediyl, preferably 2,5-tetrahydrofuranediyl or 2,5-dioxanediyl. A particular heterocyclyl group considered is a saccharide moiety, e.g. an α- or β-furanosyl or α- or β-pyranosyl moiety.

Cyclic substructures in a linker R_2 or R_4 reduce the molecular flexibility as measured by the number of rotatable bonds within R_2 or R_4, which leads to a better membrane permeation rate, important for all *in vivo* labeling applications.

A linker R_2 or R_4 is preferably a straight chain alkylene group with 1 to 25 carbon atoms or a straight chain polyethylene glycol group with 4 to 100 ethyleneoxy units, optionally attached to the group A or R_3, respectively, by a –CH=CH– or –C≡C– group. Further preferred is a straight chain alkylene group with 1 to 25 carbon atoms wherein carbon atoms are optionally replaced by an amide function –NH–CO–, and carrying a photo-cleavable subunit, e.g. o-nitrophenyl. Further preferred are branched linkers comprising a polyethylene glycol group of 3 to 6 ethylene glycol units and alkylene groups wherein carbon atoms are replaced by amide bonds, and further carrying substituted amino and hydroxy functions. Other preferred branched linkers have dendritic (tree-like) structures wherein amine, carboxamide and/or ether functions replace carbon atoms of an alkylene group.

A particularly preferred linker R_2 or R_4 is a straight chain alkylene group of 10 to 40 carbon atoms wherein 3 to 12 carbon atoms are replaced by oxygen, one or two carbon atoms are replaced by one or two 1,4-triazolidene units, respectively, and optionally one carbon atom is replaced by a 1,4-phenylene unit.

Another particularly preferred linker R_2 or R_4 is a straight chain alkylene group of 10 to 40 carbon atoms optionally substituted by oxo wherein 3 to 12 carbon atoms are replaced by oxygen and one or two carbon atoms are replaced by nitrogen.

Another particularly preferred linker R_2 or R_4 is a straight chain alkylene group of 6 to 40 carbon atoms wherein 2 to 12 carbon atoms are replaced by oxygen and one or two bonds between two adjacent carbon atoms is a double bond representing a function –CH=CH–.

Another particularly preferred linker R_2 or R_4 is a straight chain alkylene group of 1 to 15 carbon atoms, and an N-methylisoxazolidine-3,5-dimethyl group.

A linker R_2 or R_4 may carry one or more same or different labels, e.g. 1 to 100 same or different labels, in particular 1 to 5, preferably one, two or three, in particular one or two same or different labels.

R_3 as an aromatic or a heteroaromatic group, or as an optionally substituted unsaturated alkyl, cycloalkyl or heterocyclyl group is a group sterically and electronically accepted by AGT (in accordance with its reaction mechanism), which allows the covalent transfer of the R_3–R_4–L_1 unit to the fusion protein. In a R_3–R_4–L_1 unit, R_4–L_1 may also have the meaning of a plurality of same or different linkers R_4 carrying a plurality of same or different labels L_1.

R_3 as an aromatic group is preferably phenyl or naphthyl, in particular phenyl, e.g. phenyl substituted by R_4 in para or meta position.

A heteroaromatic group R_3 is a mono- or bicyclic heteroaryl group comprising zero, one, two, three or four ring nitrogen atoms and zero or one oxygen atom and zero or one sulfur atom, with the proviso that at least one ring carbon atom is replaced by a nitrogen, oxygen or sulfur atom, and which has 5 to 12, preferably 5 or 6 ring atoms; and which in addition to carrying a substituent R_4 may be unsubstituted or substituted by one or more, especially one, further substituents selected from the group consisting of lower alkyl, such as methyl, lower alkoxy, such as methoxy or ethoxy, halogen, e.g. chlorine, bromine or fluorine, halogenated lower alkyl, such as trifluoromethyl, or hydroxy.

Preferably the mono- or bicyclic heteroaryl group R_3 is selected from 2H-pyrrolyl, pyrrolyl, imidazolyl, benzimidazolyl, pyrazolyl, indazolyl, purinyl, pyridyl, pyrazinyl, pyrimidinyl, pyridazinyl, 4H-quinolizinyl, isoquinolyl, quinolyl, phthalazinyl, naphthyridinyl, quinoxalyl, quinazolinyl, quinolinyl, pteridinyl, indolizinyl, 3H-indolyl, indolyl, isoindolyl, oxazolyl, isoxazolyl, thiazolyl, isothiazolyl, triazolyl, tetrazolyl, furazanyl, benzo[d]-pyrazolyl, thienyl, and furanyl. More preferably the mono- or bicyclic heteroaryl group is selected from the group consisting of pyrrolyl, imidazolyl, such as 1H-imidazol-1-yl, benzimidazolyl, such as 1-benzimidazolyl, indazolyl, especially 5-indazolyl, pyridyl, e.g. 2-, 3- or 4-pyridyl, pyrimidinyl, especially 2-pyrimidinyl, pyrazinyl, isoquinolinyl, especially 3-isoquinolinyl, quinolinyl, especially 4- or 8-quinolinyl, indolyl, especially 3-indolyl, thiazolyl, triazolyl, tetrazolyl, benzo[d]pyrazolyl, thienyl, and furanyl.

In a particularly preferred embodiment of the invention the heteroaryl group R_3 is triazolyl, especially 1-triazolyl, carrying the further substituent R_4 in the 4- or 5-position, tetrazolyl, especially 1-tetrazolyl, carrying the further substituent R_4 in the 4- or 5-position, or 2-tetrazolyl carrying the further substituent R_4 in 5-position, isoxazolyl, especially 3-isoxazolyl carrying the further substituent R_4 in 5-position, or 5-isoxazolyl, carrying the further substituent R_4 in 3-position, or thienyl, especially 2-thienyl, carrying the further substituent R_4 in 3-, 4- or 5-position, preferably 4-position, or 3-thienyl, carrying the further substituent R_4 in 4-position.

Most preferred is the heteroaryl group R_3 as triazolyl, carrying the substituent R_4 in 4- or 5- position, and also R_3 as 2-thienyl carrying the substituent R_4 in 4- or 5-position.

An optionally substituted unsaturated alkyl group R_3 is 1-alkenyl carrying the further substituent R_4 in 1- or 2-position, preferably in 2-position, or 1-alkynyl. Substituents considered in 1-alkenyl are e.g. lower alkyl, e.g. methyl, lower alkoxy, e.g. methoxy, lower acyloxy, e.g. acetoxy, or halogenyl, e.g. chloro. In a particularly preferred embodiment of the invention R_3 is 1-alkynyl.

An optionally substituted unsaturated cycloalkyl group is a cycloalkenyl group with 5 to 7 carbon atoms unsaturated in 1-position, e.g. 1-cyclopentenyl or 1-cyclohexenyl, carrying the further substituent R_4 in any position. Substituents considered are e.g. lower alkyl,

e.g. methyl, lower alkoxy, e.g. methoxy, lower acyloxy, e.g. acetoxy, or halogenyl, e.g. chloro.

An optionally substituted unsaturated heterocyclyl group has 3 to 12 atoms, 1 to 5 heteroatoms selected from nitrogen, oxygen and sulfur, and a double bond in the position connecting the heterocyclyl group to methylene CH_2. Substituents considered are e.g. lower alkyl, e.g. methyl, lower alkoxy, e.g. methoxy, lower acyloxy, e.g. acetoxy, or halogenyl, e.g. chloro.

In particular, an optionally substituted unsaturated heterocyclyl group is a partially saturated heteroaromatic group as defined hereinbefore for a heteroaromatic group R_3. An example of such a heterocyclyl group is isoxazolidinyl, especially 3-isoxazolidinyl carrying the further substituent in 5-position, or 5-isoxazolidinyl, carrying the further substituent in 3-position.

In R_5 with the meaning arylmethyl, aryl is preferably phenyl or naphthyl, in particular phenyl or substituted phenyl, e.g. phenyl substituted in para or meta position by lower alkyl, such as methyl or ethyl, lower alkoxy, such as methoxy, halogen, e.g. fluorine or chlorine, amino, or acylamino.

In R_5 with the meaning heteroarylmethyl, heteroaryl is a mono- or bicyclic heteroaryl group comprising zero, one, two, three or four ring nitrogen atoms and zero or one oxygen atom and zero or one sulfur atom, with the proviso that at least one ring carbon atom is replaced by a nitrogen, oxygen or sulfur atom, and which has 5 to 12, preferably 5 or 6 ring atoms; and which may be unsubstituted or substituted by one or more, especially one, further substituents selected from the group consisting of lower alkyl, such as methyl, lower alkoxy, such as methoxy or ethoxy, halogen, e.g. chlorine, bromine or fluorine, halogenated lower alkyl, such as trifluoromethyl, or hydroxy. Preferred as heteroaryl in heteroarylmethyl R_5 is heteroaryl which is described as preferred under heteroaryl R_3, e.g. triazolyl or 2-thienyl.

R_5 as optionally substituted cycloalkyl is a cycloalkyl group with 3 to 7 carbon atoms, e.g. cyclopropyl, cyclopentyl or cyclohexyl, carrying an optional substituent in any position.

Substituents considered are e.g. lower alkyl, e.g. methyl, lower alkoxy, e.g. methoxy, lower acyloxy, e.g. acetoxy, or halogenyl, e.g. chloro.

R_5 as optionally substituted cycloalkenyl is a cycloalkenyl group with 5 to 7 carbon atoms unsaturated in any position, e.g. in 1-position, e.g. 1-cyclopentenyl or 1-cyclohexenyl, carrying an optional substituent in any position. Substituents considered are those listed under cycloalkyl R_5.

R_5 as optionally substituted heterocyclyl group is saturated or unsaturated and has 3 to 12 atoms, and 1 to 5 heteroatoms selected from nitrogen, oxygen and sulfur. Substituents considered are e.g. lower alkyl, e.g. methyl, lower alkenyl, e.g. vinyl or allyl, alkinyl, e.g. acetylenyl, aryl, e.g. phenyl, halo-lower alkyl, e.g. trifluoromethyl, hydroxyalkyl, e.g. hydroxymethyl, hydroxy, lower alkoxy, e.g. methoxy, lower acyloxy, e.g. acetoxy, carboxy, carbamoyl, lower alkoxycarbonyl, e.g. methoxycarbonyl, amino, acylamino, e.g. acetylamino, nitro, azido, cyano, lower alkyl- or amino-lower alkyl-sulfenyl, -sulfinyl or -sulfonyl, or halogenyl, e.g. chloro. Preferred heterocyclyl is e.g. tetrahydrofuranyl, e.g. 2-tetrahydrofuranyl.

R_5 as defined hereinbefore might have one ore more chiral centres, e.g. as in 2-tetrahydrofuranyl, leading to a selective recognition of only one enantiomer (or diastereomer) by wild type hAGT.

Preferred R_5 is cyclopentyl, cyclohexyl, benzyl, hydroxy-substituted cyclopentyl, cyclohexyl or benzyl, and hydroxy- or hydroxy-lower alkyl substituted tetrahydrofuranyl.

The labels L_1 and L_2 of the substrate can be chosen by those skilled in the art dependent on the application for which the fusion protein is intended. Labels may be e.g. such that the labelled fusion protein carrying label L_1 is easily detected or separated from its environment. Other labels considered are those which are capable of sensing and inducing changes in the environment of the labelled fusion protein and/or the substrate, or labels which aid in manipulating the fusion protein by the physical and/or chemical properties of the substrate and specifically introduced into the fusion protein.

Examples of labels L_1 and L_2 include a spectroscopic probe such as a fluorophore or a chromophore, a magnetic probe or a contrast reagent; a radioactively labelled molecule; a molecule which is one part of a specific binding pair which is capable of specifically binding to a partner; a molecule that is suspected to interact with other biomolecules; a library of molecules that are suspected to interact with other biomolecules; a molecule which is capable of crosslinking to other molecules; a molecule which is capable of generating hydroxyl radicals upon exposure to H_2O_2 and ascorbate, such as a tethered metal-chelate; a molecule which is capable of generating reactive radicals upon irradiation with light, such as malachite green; a molecule covalently attached to a solid support, where the support may be a glass slide, a microtiter plate or any polymer known to those proficient in the art; a nucleic acid or a derivative thereof capable of undergoing base-pairing with its complementary strand; a lipid or other hydrophobic molecule with membrane-inserting properties; a biomolecule with desirable enzymatic, chemical or physical properties; or a molecule possessing a combination of any of the properties listed above. Preferred are labels L_1 and L_2 as mentioned hereinbefore with the exception of a radioactively labelled molecule. Excluded from the scope are labels L_2 with the meaning of a nucleic acid. Most preferred as labels L_1 are spectroscopic probes, and molecules which are one part of a specific binding pair which is capable of specifically binding to a partner, so-called affinity labels. Most preferred as labels L_2 are molecules representing one part of a specific binding pair which is capable of specifically binding to a partner, and molecules covalently attached to a solid support.

When the label L_1 or L_2 is a fluorophore, a chromophore, a magnetic label, a radioactive label or the like, detection is by standard means adapted to the label and whether the method is used *in vitro* or *in vivo*. If L_1 is a fluorophore the method can be compared to the applications of the green fluorescent protein (GFP) which is genetically fused to a protein of interest and allows protein investigation in the living cell. Particular examples of labels L_1 and L_2 are also boron compounds displaying non-linear optical properties. Particularly preferred are labels such that L_1 is one member and L_2 is the other member of two interacting spectroscopic probes L_1 / L_2, wherin energy can be transferred nonradiatively between the donor and acceptor (quencher) when they are in close proximity (less than 1 nanometer distance) through either dynamic or static quenching. Such a pair of labels L_1 / L_2 changes its spectroscopic properties on reaction of the

labelled substrate with the AGT fusion protein. An example of such a pair of labels L_1 / L_2 is a FRET pair explained below in more detail.

Depending on the properties of the label L_1, the fusion protein comprising protein of interest and AGT may be bound to a solid support on reaction with the substrate. The label L_1 of the substrate reacting with the fusion protein comprising AGT may already be attached to a solid support when entering into reaction with AGT, or may subsequently, i.e. after transfer to AGT, be used to attach the labelled AGT fusion protein to a solid support. Alternatively, the label L_2 of the substrate may be a solid support or attached or attachable to a solid support, which allows separating the labelled fusion protein carrying label L_1 from the remaining part of the substrate after reaction and containing L_2. The label may be one member of a specific binding pair, the other member of which is attached or attachable to the solid support, either covalently or by any other means. A specific binding pair considered is e.g. biotin and avidin or streptavidin. Either member of the binding pair may be the label L_1 and/or L_2 of the substrate, the other being attached to the solid support. Further examples of labels allowing convenient binding to a solid support are e.g. maltose binding protein, glycoproteins, FLAG tags, or reactive substituents allowing chemoselective reaction between such substituent with a complementary functional group on the surface of the solid support. Examples of such pairs of reactive substituents and complementary functional group are e.g. amine and activated carboxy group forming an amide, azide and a propiolic acid derivative undergoing a 1,3-dipolar cycloaddition reaction, amine and another amine functional group reacting with an added bifunctional linker reagent of the type of activated bis-dicarboxylic acid derivative giving rise to two amide bonds, or other combinations known in the art.

Examples of a convenient solid support are e.g. glass surfaces such as glass slides, microtiter plates, and suitable sensor elements, in particular functionalized polymers (e.g. in the form of beads), chemically modified oxidic surfaces, e.g. silicon dioxide, tantalum pentoxide or titanium dioxide, or also chemically modified metal surfaces, e.g. noble metal surfaces such as gold or silver surfaces. Irreversibly attaching and/or spotting AGT substrates may then be used to attach AGT fusion proteins in a spatially resolved manner, particularly through spotting, on the solid support representing protein microarrays, DNA microarrays or arrays of small molecules.

When the label L_1 or L_2 is capable of generating reactive radicals, such as hydroxyl radicals, upon exposure to an external stimulus, the generated radicals can then inactivate the AGT fusion proteins as well as those proteins that are in close proximity of the AGT fusion protein, allowing to study the role of these proteins. Examples of such labels are tethered metal-chelate complexes that produce hydroxyl radicals upon exposure to H_2O_2 and ascorbate, and chromophores such as malachite green that produce hydroxyl radicals upon laser irradiation. The use of chromophores and lasers to generate hydroxyl radicals is also known in the art as chromophore assisted laser induced inactivation (CALI). In the present invention, labeling AGT fusion proteins with substrates carrying chromophores as label L_1, such as malachite green, and subsequent laser irradiation inactivates the labelled AGT fusion protein as well as those proteins that interact with the AGT fusion protein in a time-controlled and spatially-resolved manner. This method can be applied both *in vivo* or *in vitro.* Furthermore, proteins which are in close proximity of the AGT fusion protein can be identified as such by either detecting fragments of that protein by a specific antibody, by the disappearance of those proteins on a high-resolution 2D-electrophoresis gels or by identification of the cleaved protein fragments *via* separation and sequencing techniques such as mass spectrometry or protein sequencing by N-terminal degradation.

When the label L_1 is a molecule that can cross-link to other proteins, e.g. a molecule containing functional groups such as maleimides, active esters or azides and others known to those proficient in the art, contacting such labelled AGT substrates with AGT fusion proteins that interact with other proteins (*in vivo* or *in vitro*) leads to the covalent cross-linking of the AGT fusion protein with its interacting protein *via* the label. This allows the identification of the protein interacting with the AGT fusion protein. Labels L_1 (and L_2) for photo cross-linking are e.g. benzophenones. In a special aspect of cross-linking the label L_1 is a molecule which is itself an AGT substrate leading to dimerization of the AGT fusion protein. The chemical structure of such dimers may be either symmetrical (homodimers) or unsymmetrical (heterodimers).

Other labels L_1 considered are for example fullerenes, boranes for neutron capture treatment, nucleotides or oligonucleotides, e.g. for self-addressing chips, peptide nucleic acids, and metal chelates, e.g. platinum chelates that bind specifically to DNA.

A particular biomolecule with desirable enzymatic, chemical or physical properties is methotrexate. Methotrexate is a tight-binding inhibitor of the enzyme dihydrofolate reductase (DHFR). Compounds of formula (1) wherein L_1 is methotrexate belong to the well known class of so-called "chemical inducers of dimerization" (CIDs). Using fusion proteins of hAGT with the DNA-binding domain LexA, and adding DHFR with the transcriptional activation domain B42 to the *in vivo* labeling of the hAGT fusion protein with a compound of formula (1) wherein L_1 is methotrexate induces the coupling ("dimerization") of the hAGT-LexA fusion protein and DHFR-B42 fusion protein, leading to spatial proximity of LexA and B42 and subsequent stimulation of transcription.

If the substrate carries two or more labels, these labels may be identical or different. Particular preferred combinations are two different affinity labels, or one affinity label and one chromophore label, in particular one affinity label and one fluorophore label, or a pair of spectroscopic interacting labels L_1 / L_2, e.g a FRET pair.

The invention further relates to a method for detecting and/or manipulating a protein of interest, wherein the protein of interest is incorporated into an AGT fusion protein, the AGT fusion protein is contacted with particular AGT substrates carrying a label described hereinbefore, and the AGT fusion protein is detected and optionally further manipulated using the label in a system designed for recognising and/or handling the label.

In the method of the present invention a protein or peptide of interest is fused to an O^6-alkylguanine-DNA alkyltransferase (AGT). The protein or peptide of interest may be of any length and both with and without secondary, tertiary or quaternary structure, and preferably consists of at least twelve amino acids and up to 2000 amino acids. Examples of such protein or peptide of interest are e.g. enzymes, DNA-binding proteins, transcription regulating proteins, membrane proteins, nuclear receptor proteins, nuclear localization signal proteins, protein cofactors, small monomeric GTPases, ATP-binding cassette proteins, intracellular structural proteins, proteins with sequences responsible for targeting proteins to particular cellular compartments, proteins generally used as labels or affinity tags, and domains or subdomains of the aforementioned proteins. The protein or peptide of interest is preferably fused to AGT by way of a linker which may be

cleaved by an enzyme, e.g. at the DNA stage by suitable restriction enzymes and/or linkers cleavable by suitable enzymes at the protein stage.

The O^6-alkylguanine-DNA alkyltransferase (AGT) has the property of transferring a label present on a substrate to one of the cysteine residues of the AGT forming part of a fusion protein. In preferred embodiments, the AGT is wild type human O^6-alkylguanine-DNA alkyltransferase, hAGT, or a mutant thereof, e.g. a mutant as described in Juillerat *et al.*, Chem Biol 10:313-317, 2003, or in the Examples hereinafter. A mutant AGT to be used in the invention may differ from wild type hAGT by virtue of one or more amino acid substitutions, deletions or additions, but still retains the property of transferring a label present on a substrate to the AGT part of the fusion protein. A mutant AGT may preferably be produced using protein engineering techniques known to the skilled person in the art, e.g. saturation mutagenesis, error prone PCR to introduce variations anywhere in the sequence, or DNA shuffling used after saturation mutagenesis and/or error prone PCR.

The fusion protein comprising protein of interest and an O^6-alkylguanine-DNA alkyl-transferase (AGT) is contacted with a particular substrate having a label. Conditions of reaction are selected such that the AGT reacts with the substrate and transfers the label of the substrate. Usual conditions are a buffer solution at around pH 7 at room temperature, e.g. around 25°C. However, it is understood that AGT reacts also under a variety of other conditions, and those conditions mentioned here are not limiting the scope of the invention.

AGT irreversibly transfers the alkyl group from its natural substrate, O^6-alkylguanine-DNA, to one of its cysteine residues. Likewise, AGT transfers the guanine type O^6 substituent or the corresponding substituent in a related position of the substrates of the invention to one of its cysteine residues. This property of AGT is used in the method of the invention to transfer the label L_1 attached to the residue CH_2–R_3–R_4– of a compound of formula (1) to AGT.

This label L_1 of the substrate can be chosen by those skilled in the art dependent on the application for which the fusion protein is intended. After contacting the fusion protein comprising AGT with the substrate, the label L_1 is covalently bonded to the fusion

protein. The labelled AGT fusion protein is then further manipulated and/or detected by virtue of the transferred label. The label L_1 may consist of a plurality of same or different labels. If the substrate contains more than one label L_1, the corresponding labelled AGT fusion protein will also comprise more than one label which gives more options for further manipulating and/or detecting the labelled fusion protein.

In a particular aspect, the present invention provides a particularly convenient method to transfer a label L_1 to the AGT fusion protein and to remove all unreacted AGT substrate. This is simply done in the case when label L_2 connected through R_2 to the heterocyclic group A recognized by AGT as a substrate is a label L_2 with the meaning of a solid support or the meaning of a reactive group attachable to a solid support, e.g. one part of a specific binding pair which is capable of specifically binding to a partner attached to a solid support. The solid support and the protein solution containing the AGT fusion protein labelled with L_1 are then easily separated, e.g. by filtration, without further manipulation. More specifically, a molecular excess of substrate of the formula (1) wherein R_1 is R_2–L_2 and L_2 is a solid support or a group attached to a solid support is reacted with an AGT fusion protein, and the obtained labelled AGT fusion protein carrying a label L_1 is separated from the solid support to which the remaining part of the reacted substrate, i.e. the residue R_2–A, is bound, and also to which the excess unreacted substrate of formula (1) remains bound. In other terms the AGT fusion protein, on reacting with the substrate, removes and takes over the label L_1 from a solid support, leaving all unreacted label L_1 bound to the solid support.

In this particular aspect the label L_1 may be any label as described hereinbefore, for example a spectroscopic probe or a fluorophore. The solid support, i.e. L_2 or the support to which L_2 is attached, may be any solid support, for example a bead (e.g. a magnetic bead), a polymer support or a metal surface. Separation of the labelled AGT fusion protein from unreacted substrate may be by filtration, centrifugation or other suitable methods depending on the properties of the solid support, e.g. applying a magnetic field if the solid support is a magnetic bead.

In another particular aspect of the invention a method is provided which makes use of a spatial separation of labels L_1 and L_2 on reaction with an AGT fusion protein. Particular substrates of the present invention are designed as interacting spectroscopic labels, e.g.

pro-fluorescent probes covalently labeled at one end (e.g. L_1) with a donor (reporter) and at the other end (L_2) with an acceptor (quencher), and *vice versa*. It is generally known that quenching in such probes may occur through Förster resonance energy transfer (FRET) or through static quenching by close proximity of the donor acceptor pair. The donor and acceptor should be chosen such as to maximize their spectral overlap. Förster energy transfer typically occurs over distances of up to 20-100 Å.

In the particular method of the invention, wherein L_1 is a donor (reporter) and L_2 is an acceptor (quencher), or wherein L_1 is a quencher and L_2 is a reporter the reaction of the AGT fusion protein with the substrate leads to a change in fluorescence. The reporter-quencher distance within the doubly labeled substrate is changed upon reaction with the AGT fusion protein leading to a spatial separation of reporter and quencher witch causes the appearance of fluorescence. A broad selection of reporter groups may be used as the label L_1 or L_2, respectively, including e.g. infra-red emitting fluorophores, and the sensitivity and specificity of the method is substantially improved compared to substrates carrying only one fluorophore label L_1. The substrate containing reporter and quencher remains dark until it reacts with the AGT fusion protein, whereupon the reaction mixture is "lit up" switching on the fluorophore emission, since the reporter label and the quencher label are now spatially separated. Fluorescence quenching and energy transfer can be measured by the emission of only one of the two labels, the quenched or energy donor label. When energy transfer occurs and the energy accepting label is also fluorescent, the acceptor label fluorescence can also be measured. A donor label of these two interacting labels can be chosen from chemiluminescent donor probes which eliminates the need of an excitation lamp and reduces acceptor background fluorescence. The mentioned particular method using such double-labelled substrates is useful to determine reaction kinetics based on fluorescence time measurements, and may be applied *in vivo* as well as *in vitro.*

In vitro, the reaction of the AGT fusion protein with the substrate of the invention can generally be either performed in cell extracts or with purified or enriched forms of the AGT fusion protein.

If experiments with the substrates of the present invention are done *in vivo* or in cell extracts, the reaction of the endogenous AGT of the host is advantageously taken into

account. If the endogenous AGT of the host does not accept O^6-alkylguanine derivatives or related compounds as a substrate, the reaction of the (exogenous) AGT fusion protein is specific. In mammalian cells, e.g. in human, murine, or rat cells, unspecific reaction with endogenous AGT is possible. In those experiments where the simultaneous reaction of the endogenous AGT as well as of the (exogenous) AGT fusion protein poses a problem, known AGT-deficient cell lines can be used.

In the described particular embodiment of the invention, the doubly labelled substrates carrying reporter and quencher allow to determine the concentration of AGT fusion proteins either *in vitro* or *in vivo*. For *in vivo* application, the reporters are preferably emitters in the near infra red (NIR) region because that region is absent of interfering bio fluorescence. Known cyanine NIR dyes matching these requirements are preferably incorporated in the substrates of the present invention.

Appropriate pairs of reporters and quenchers can bee chosen by those skilled in the art. Typically reporter and quencher are fluorescent dyes with large spectral overlap as, for example, fluorescein as a reporter and rhodamine as a quencher. Other quenchers are gold clusters, and metal cryptates.

A second class of quenchers used in this invention are "dark quenchers" (Johnasson, M.K. *et al.*, Chem. Eur. J. 9:3466-3471, 2003), i.e. dyes without native fluorescence having absorption spectra that overlap with the emission spectra of common reporter dyes leading to maximal FRET quenching. Furthermore pairs of dyes can be chosen such that their absorption bands overlap in order to promote a resonance dipole-dipole interaction mechanism within a ground state complex (static quenching).

Particular pairs of fluorophores and quenchers considered are: Alexa dyes (Panchuk-Voloshina, N. *et al.*, J. Histochem. & Cytochem. 47:1179-1188, 1999) and quenchers QSY 7, QSY 9, QSY 21, or QSY 35 (Molecular Probes, Inc., Eugene, OR 97402, USA); Cy 3.5 or 6-carboxyfluorescein and BHQ1 (Black Hole Quencher™ of Biosearch Technologies, Inc., Novato, CA 94949, USA); and fluorescein, Alexa 350, tetratemethyl-rhodamine or Cyanine-5 (Cy 5) and 4-dimethylaminoazobenzene-4'-sulfonyl chloride (Dabsyl).

In a particular embodiment, the method involves a substrate wherein L_2 is a solid support or attached to a solid support further carrying one member of the reporter / quencher pair, or wherein L_2 is a combination of a solid support and one member of the reporter / quencher pair, and L_1 is the other member of this pair. In this way, the dark solid support being an AGT substrate becomes fluorescent upon reaction with the AGT fusion protein.

The mentioned particular method involving fluorescence determination based on a reporter / quencher pair, e.g. a FRET pair, is much more convenient than other methods for measuring the concentration or the kinetics of reactivity based on SDS/PAGE, Western blotting or the like, which require considerable protein preparation and manipulation of conditions to remove non-specific signals. The energy transfer method does not require separation of reacted and unreacted AGT substrate. Therefore this method is particularly convenient for high-throughput screening of compound libraries, identification of active AGT mutants or identification of AGT inhibitors.

In a further particular method of the invention, a substrate is used wherein R_1 is a group R_5, in particular a cycloalkyl, cycloalkenyl, cycloalkylmethyl, arylmethyl or heteroaryl-methyl group. Examples of preferred groups R_5 are cyclopentyl and benzyl. In the AGT fusion protein, a mutant AGT is chosen as fusion partner which has a low reactivity towards substrates wherein R_5 is one of the mentioned particular or preferred substituents. In such a particular method for detecting and/or manipulating a protein of interest, the protein of interest is fused with the mutant AGT, the mutant AGT fusion protein is contacted with a mixture of (a) a substrate wherein R_1 is a group R_5 and which is not recognized by the mutant AGT, and (b) another AGT substrate recognized by the mutant AGT fusion protein, and carrying a label described hereinbefore, and the mutant AGT fusion protein is detected and optionally further manipulated using the label in a system designed for recognizing and/or handling the label. Such a method is the method of choice if the reaction is carried out *in vivo* in a system comprising endogenous AGT, or carried out in cell extracts comprising endogenous AGT. The endogenous AGT is saturated, and if required, separated using the label L_1 of the substrate of the invention wherein R_1 is a group R_5. In a parallel reaction the mutant AGT fusion protein reacts with another AGT substrate carrying a label or a plurality of labels as described hereinbefore, and detected and/or manipulated as required.

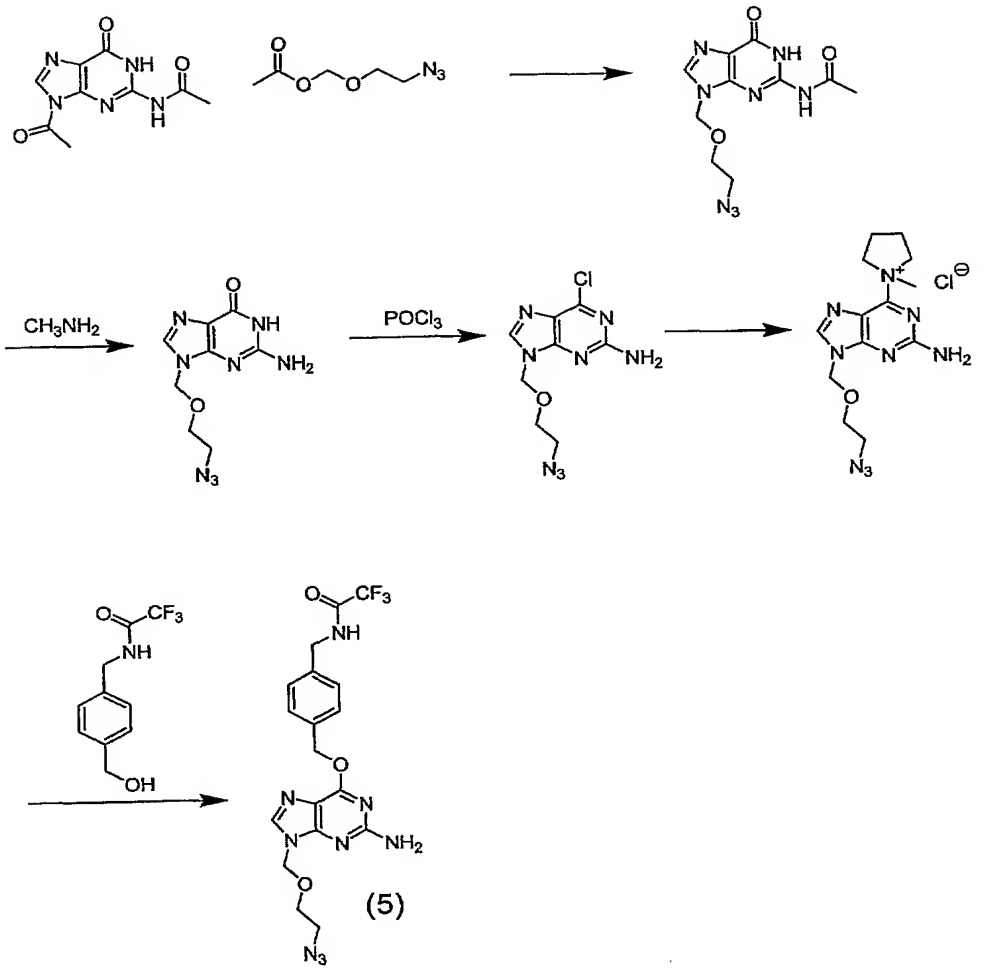
Method of manufacture

Substrates of the invention are generally prepared by standard methods known in the art. Particular methods are explained e.g. in patent application PCT/EP03/10889. For the synthesis of a substrate of formula (2), i.e. compounds wherein A is a purine entity, known methods can be used to modify the O^6, C^8 and N^9 position. The present invention also relates to new methods as described hereinafter, and to novel intermediates used and obtained.

In particular, the preferred synthesis of a AGT substrate carrying a plurality of labels (e.g. L_1 and L_2) makes use of orthogonally protected functional groups. Such a choice of protective groups allows for a separate deprotection so that each released functionality in turn can be further chemically manipulated either to attach a label to it or for the introduction of further extension of the linker R_2 and/or R_4. Appropriate protecting groups for the envisioned functionalities can be chosen by those skilled in the art, and are e.g. summarized in T.W. Greene and P.G.M. Wuts in "Protective Groups in Organic Synthesis", John Wiley & Sons, New York 1991.

For the preparation of guanine derivatives, the group $-CH_2-R_3-R_4-L_1$ or a precursor thereof is introduced on reaction of a corresponding 6-chloroguanine derivative *via* the intermediacy of an ammonium salt with a tertiary amine, e.g. methylpyrrolidine. Displacement of the ammonium group by an alcoholate $^-O-CH_2-R_3-R_4-L_1$ or a precursor thereof gives higher overall yields compared to the direct displacement of chlorine, and the reaction with ammonium salts can be performed at room temperature.

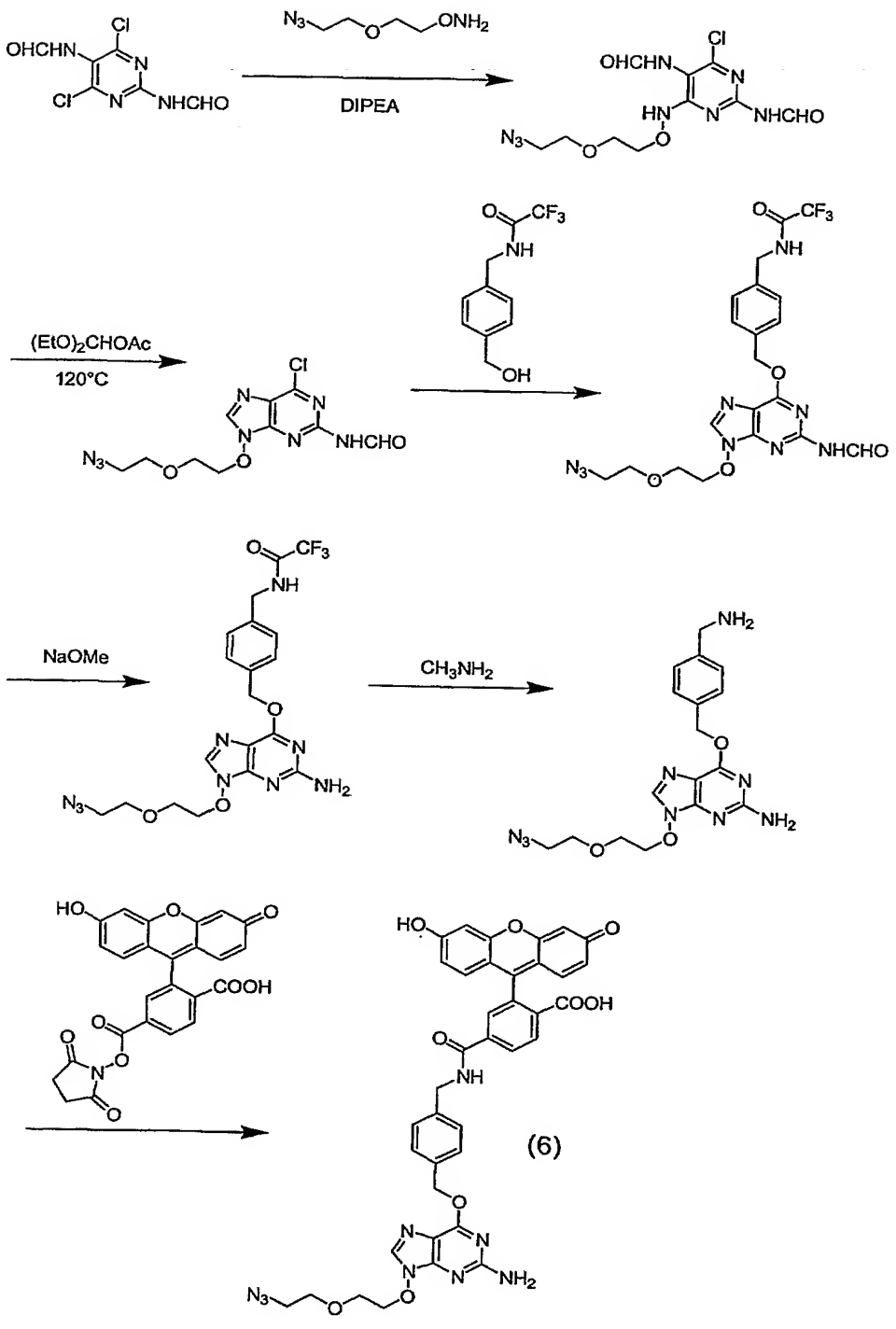
Introduction of a group R_1 at position N^9 of a guanine of formula (2) may be performed by direct alkylation *via* SN_2 reaction using halogen derivatives R_1–Hal. However, regioselectivity is generally poor, and a product mixture of N^7 and N^9 alkylated guanines is obtained. A higher N^9/N^7 ratio in the alkylated product is obtained on activating the imidazole ring by pre-treatment with LiH, or by equilibrating the obtained product at higher temperatures which favors the N^9 substitution. The preparation of a convenient intermediate of formula (5) is shown in Scheme 1:



Scheme 1

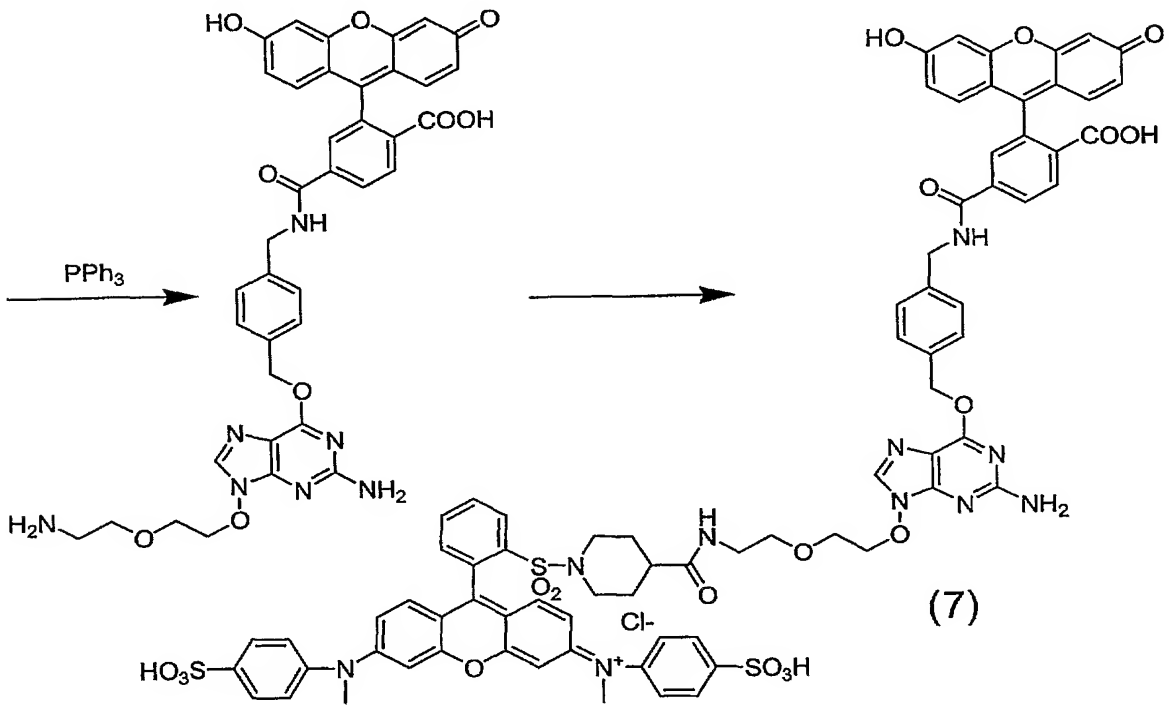
The intermediate of formula (5) is orthogonally protected at the linker in N^9 and at the benzyl function in position O^6. This intermediate may then be further manipulated to complete the residue $-R_4-L_1$ at the benzyl function in O^6 and the residue $-R_2-L_2$ at the position N^9. Alternatively Mitsunobu conditions (alcohol derivative R_1-OH, triphenylphosphine, diethyl azodicarboxylate) may be used to introduce a substituent R_1 (i.e. R_5 or $-R_2-L_2$) in position N^9.

Scheme 2 demonstrates a ring closure reaction leading to regioselective introduction of a substituent in position N^9. Intermediate of formula (6) already contains a label L_1 and allows the introduction of a suitable label L_2 by modification of the azido function.



Scheme 2

Reduction of the azido function in the compound of formula (6) with triphenylphosphine and acylation with QSY-9-NHS (Molecular Probes) gives the pro-fluorescent AGT substrate of formula (7), wherein L_1 (carboxy-fluorescein) and L_2 (QSY-9) represent a FRET pair (Scheme 3).



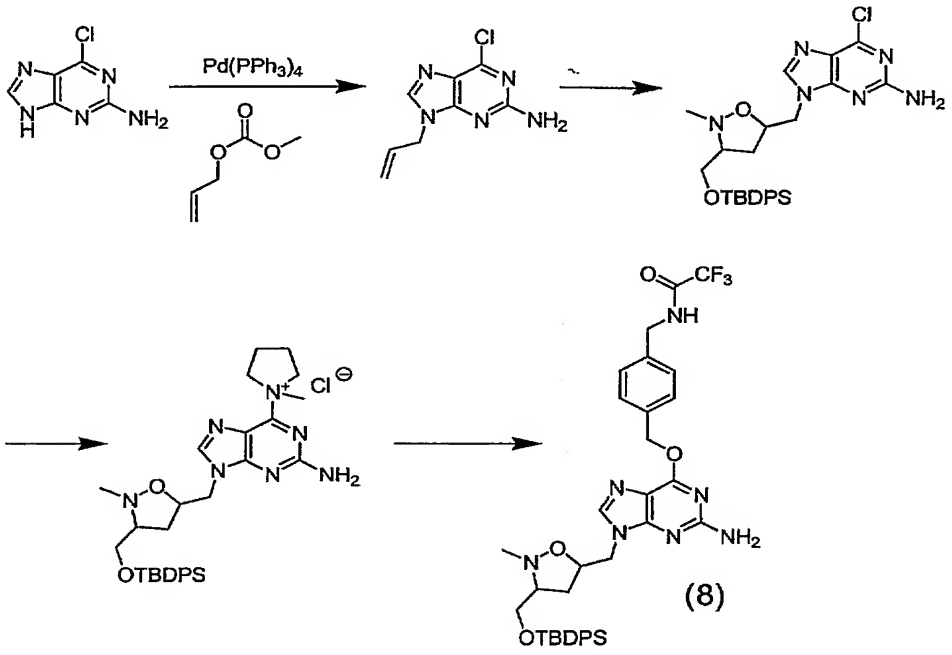
Scheme 3

A further method for regioselective alkylation in N^9 is palladium catalyzed allylic alkylation as shown in Scheme 4.

For structural modifications at the level of the linker moiety R_2, the synthesis of appropriate functionalized homo-N,O-nucleosides, wherein the sugar moiety of guanosine is replaced by an isoxazolidine ring, is likewise shown in Scheme 4 (TBDPS = tert-butyl-diphenyl-silyl). Such modifications introduce an increased resistance to hydrolytic or enzymatic cleavage compared to the relatively reactive aminal linkage of common nucleosides and more conformational flexibility. The reaction sequence is

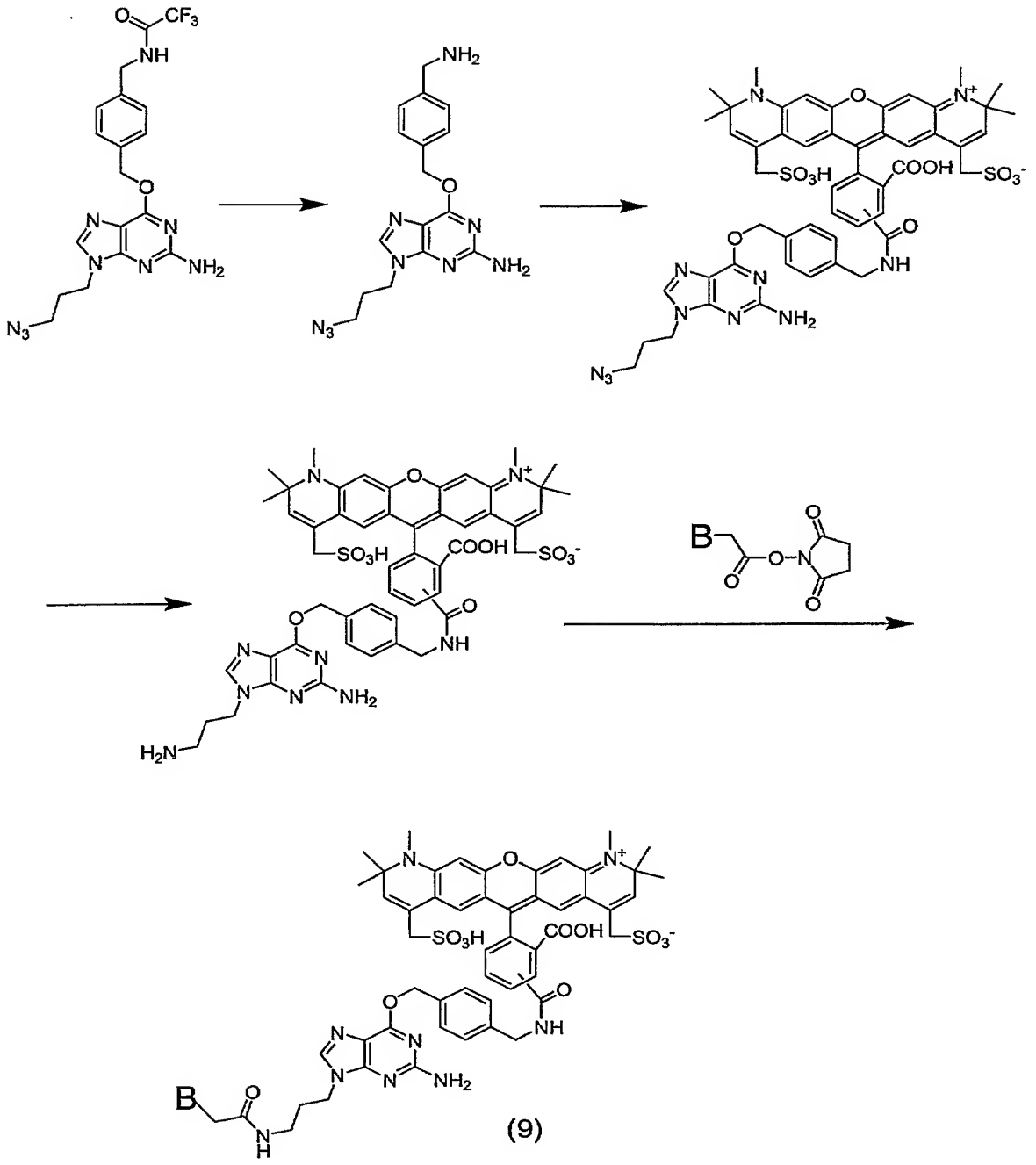
versatile, allowing for the attachment of different appropriate functionalized nitrones to allyl double bond.

Alternatively, the allyl group at N^9 may be transformed into a primary hydroxy function with the boron hydride reagent 9-borabicyclo[3.3.1]nonane (9-BBN), and then further manipulated to give appropriate groups R_5 or $-R_2-L_2$.



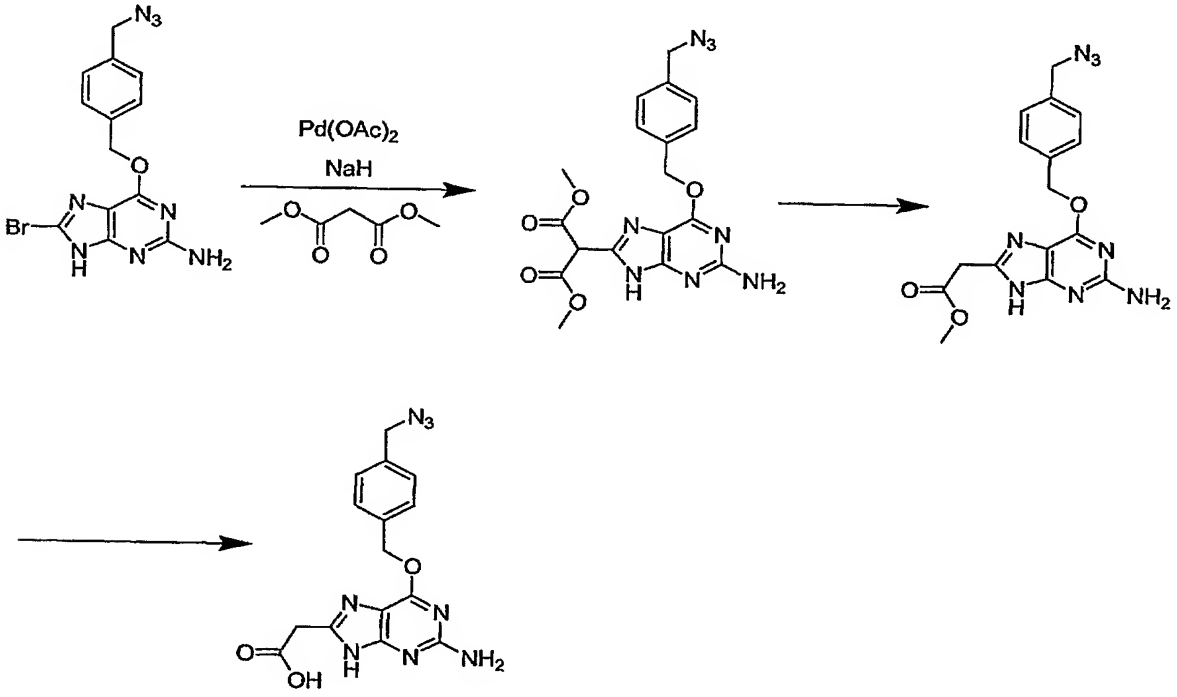
Scheme 4

The synthesis of a substrate of the invention of formula (9) wherein L_1 is a fluorophore (Alexa 594, Molecular Probes) and L_2 is a solid support (e.g. a bead B, or a biotin residue B attachable to a bead carrying avidin or streptavidin) is summarized in Scheme 5.



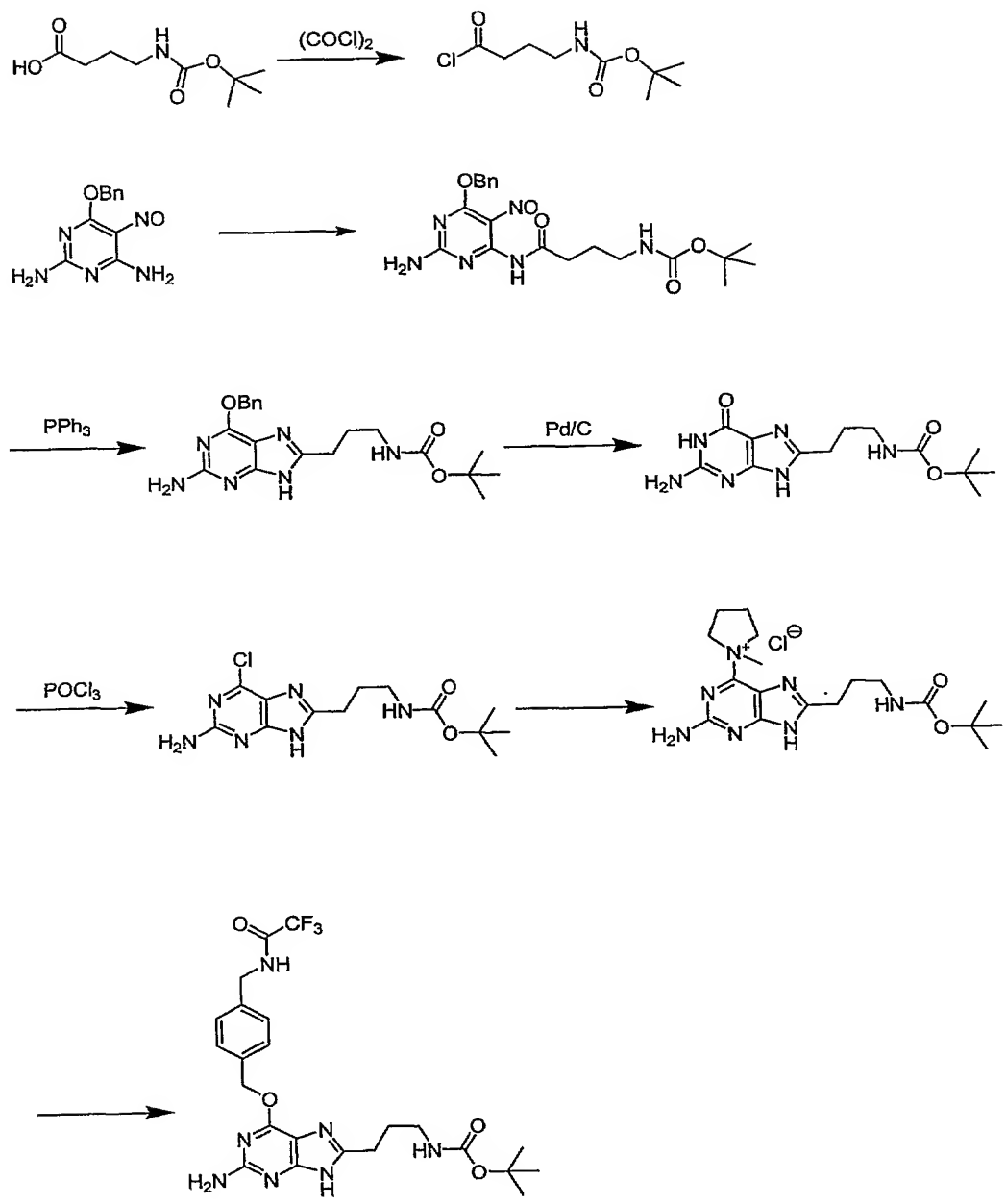
Scheme 5

The synthesis of intermediates useful in the synthesis of a compound of formula (1) wherein two different labels are attached in O^6 and C^8 position of the purine base is summarized in Schemes 6 and 7.



(10)

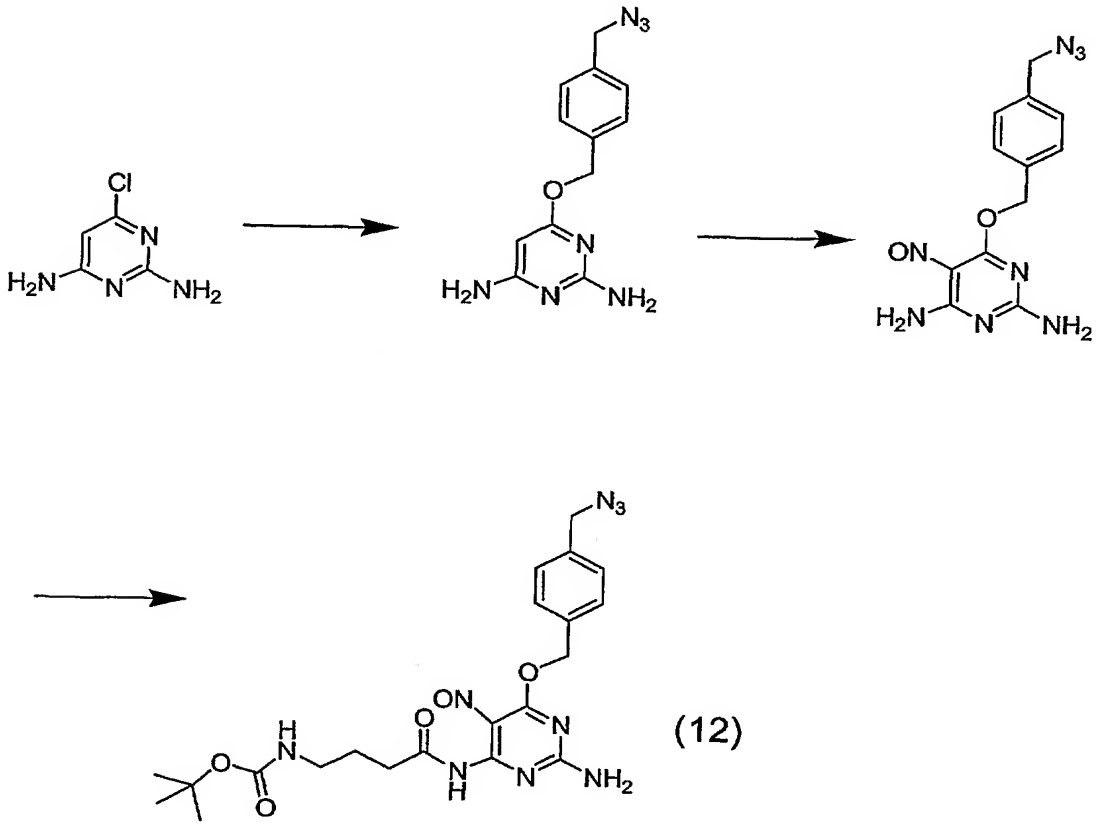
Scheme 6



Scheme 7

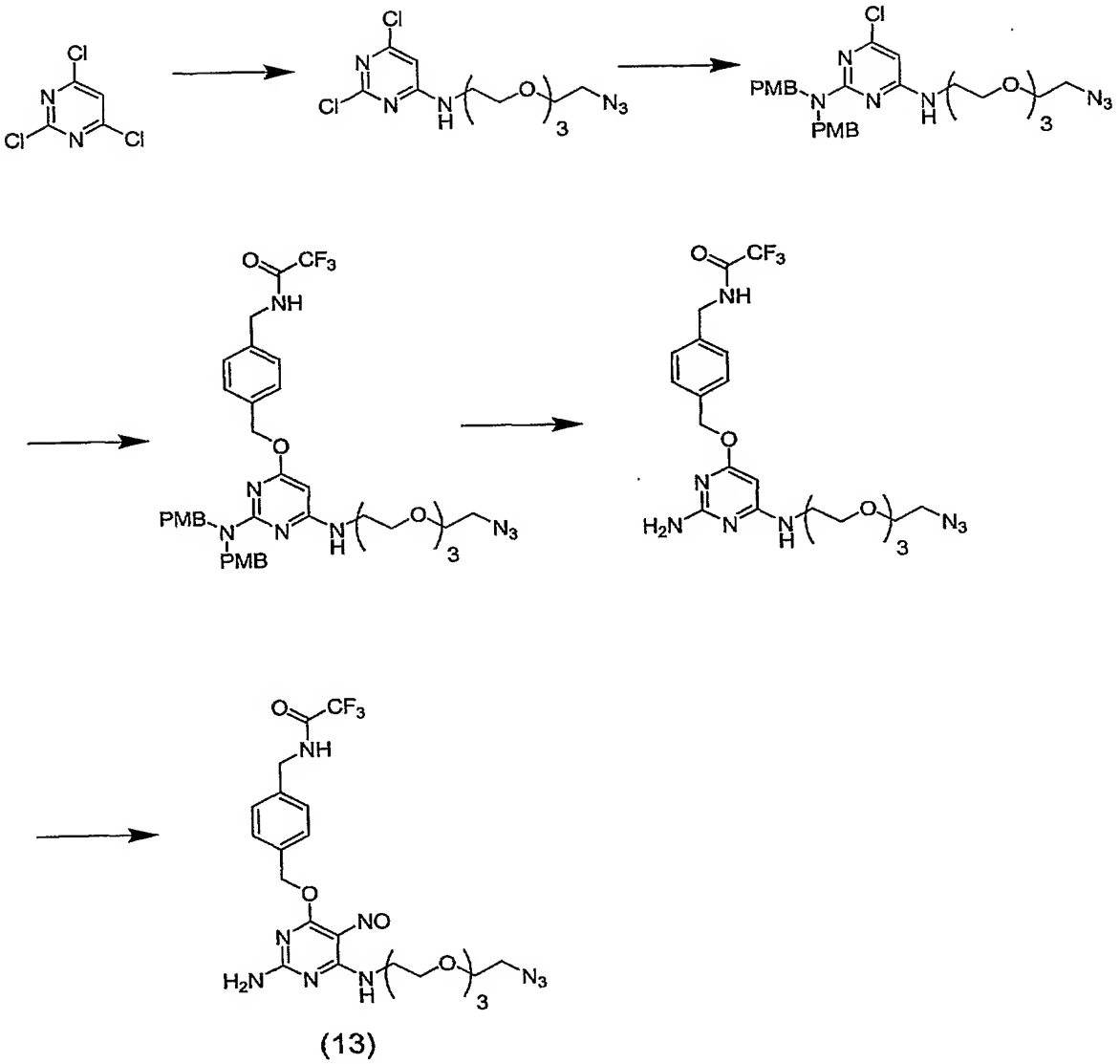
For the preparation of compounds of formula (1) wherein R_1-A is a pyrimidine residue of formula (4a), R_9 is amino and R_{10} is nitroso, the sequence of reaction shown in Scheme 8 or 9 is performed. 4-Chloro-2,6-diaminopyrimidine is reacted with the appropriate

sodium alkoxide and then treated with sodium nitrite in 30% acetic acid to introduce the nitroso function. N^6-substituted derivatives are prepared by treatment with the appropriate anhydride or acyl chloride.



Scheme 8

The introduction of an aminoalkyl substituent at the N^6 position is shown in Scheme 9. Reaction of 11-azido-3,6,9-trioxa-undeca-amine with 2,4,6-trichloropyrimidine yields the 6-substituted chloropyrimidine accompanied by the 2,6-disustituted product which can be separated by chromatography. Substitution of the 2-position with bis-p-methoxy-benzylamine (PMB) as a masked ammonia equivalent gives the disubstituted pyrimidine which is subsequently converted into the O^4 derivative by reaction with the appropriate alkoxide. Deprotection with trifluoroacetic acid at 60°C liberates the 2-aminopyrimidine. The 5-nitroso compound is prepared with sodium nitrite in acetic acid.



(13)

Scheme 9

Examples

Abbreviations:

DMSO = dimethylsulfoxide

sat. = saturated

TLC = thin layer chromatography

Example 1: (4-Bromothiophen-2-yl)-methanol (14)

$NaBH_4$ (1.11 g, 29.31 mmol) is added to a solution of 4-bromothiophene-2-carbaldehyde (5 g, 26.17 mmol) in isopropanol (70 mL). The reaction mixture is stirred for 2 h at room temperature. A saturated aqueous solution of NH_4Cl (15 mL) is added to the solution. The suspension is filtered, the filtrate concentrated *in vacuo*, dissolved in CH_2Cl_2 (70 mL), dried over $MgSO_4$ and concentrated *in vacuo* again. The residue is purified by flash chromatography (ethyl acetate / petrol ether 1 : 10). Yield 4.55 g (23.55 mmol, 90%). TLC: R_f = 0.75 (ethyl acetate / petrol ether 1 : 1). ^1H NMR ($CDCl_3$) : δ = 2.07 (broad s, 1H, -OH), 4.78 (d, 2H, C-CH_2-OH), 6.92 (s, 1H, C-CH-C(Br)), 7.17 (s, 1H, C(Br)-CH-S) ppm.

Example 2: 2-Cyclopentenyl methyl carbonate (15)

Methyl chloroformate (2.94 mL, 38 mmol) is added dropwise over 30 min. to a stirred solution of cyclopent-2-en-1-ol (1.0 g, 11.9 mmol, J.-L. Luche, J. Am. Chem. Soc., 1978, 100:7, 2226-2227) in CH_2Cl_2 (30 mL) and pyridine (10 mL) at 0°C. After 4 h TLC shows complete reaction, and the reaction mixture is poured into sat. NH_4Cl (50 mL) and extracted with Et_2O (3 x 50 mL). The organic phase is washed with HCl 1 M until the washings are acidic, washed with water (50 mL), brine (50 mL) and dried over $MgSO_4$. The crude product is purified by flash chromatography (ethyl acetate / petrol ether 1 : 100). Yield 1.777 g (12.5 mmol, 53%). TLC R_f = 0.70 (ethyl acetate / petrol ether 1 : 12). ^1H NMR ($CDCl_3$) : δ = 1.88-1.96 (m, 1H, CH_2-CH_2-CH(OH) *syn*), 2.24-2.37 (m, 2H, CH-CH_2-CH_2 *anti*, CH_2-CH_2-CH(OH) *anti*), 2.49-2.59 (m, 1H, CH-CH_2-CH_2 *syn*), 3.77 (s, 3H, O-CH_3), 5.61-5.64 (m, 1H, CH-CH(O)-CH_2), 5.87-5.89 (m, 1H, CH_2-CH=CH), 6.13-6.16 (m, 1H, CH=CH-CH(O)) ppm.

Example 3: N^9-(Cyclopent-2-enyl)-6-chloroguanine (16)

This reaction is performed under anhydrous conditions (argon atmosphere and solvents over molecular sieve). To a solution of 6-chloroguanine (596 mg, 3.52 mmol) in DMSO (10 mL), $Pd(PPh_3)_4$ (404 mg, 0.35 mmol) is added followed by a solution of 2-cyclopentenyl methyl carbonate (13) (500 mg, 3.52 mmol) in THF (10 mL). The reaction mixture is stirred 1 h at room temperature. It is poured into H_2O (60 mL) and extracted with ethyl acetate (3 x 50 mL). The combined organic layers are washed with brine (60 mL), dried over $MgSO_4$ and concentrated *in vacuo*. The residue is purified by flash

column chromatography (ethyl acetate / petrol ether 6 : 4). Yield: 381 mg (1.62 mmol, 61%). TLC R_f = 0.33 (ethyl acetate / petrol ether 6 : 4). ^{1}H NMR ($CDCl_3$) : δ = 1.88-1.95 (m, 1H, CH_2-C$\underline{H}_2$-CH(OH) *syn*), 2.49-2.71 (m, 3H, CH-C$\underline{H}_2$-CH_2 *syn*, CH-C$\underline{H}_2$-CH_2 *anti*, CH_2-C$\underline{H}_2$-CH(OH) *anti*), 5.04 (broad s, 2H, -N$\underline{H}_2$), 5.56-5.60 (m, 1H, CH-C$\underline{H}$(N)-CH_2), 5.85-5.88 (m, 1H, CH_2-C$\underline{H}$=CH), 6.29-6.32 (m, 1H, CH=C$\underline{H}$-CH(N)), 7.72 (s, 1H, N-C$\underline{H}$=N) ppm.

Example 4: N^9-Cyclopentyl-6-chloroguanine (17)

Palladium on active charcoal (100 mg) is added to a solution of N^9-(cyclopent-2-enyl)-6-chloroguanine (16) (200 mg, 0.85 mmol) in methanol (33 mL). Hydrogen gas is passed through the solution during 30 min. The crude product is adsorbed on silica (500 mg) and purified by flash chromatography (ethyl acetate / petrol ether 1 : 4, 3 : 7 and 2 : 3). Yield: 111 mg (0.47 mmol, 55%). TLC R_f = 0.34 (ethyl acetate / petrol ether 1 : 1). ^{1}H NMR ($CDCl_3$) : δ 1.79 (m, 2H, CH_2-C$\underline{H}_2$-CH(OH) *syn*), 1.94 (m, 4H, CH_2-C$\underline{H}_2$-CH_2 *anti*, CH_2-C$\underline{H}_2$-CH(OH) *anti*), 2.24 (m, 2H, CH_2-C$\underline{H}_2$-CH_2 *syn*), 4.77 (qnt, 1H, CH_2-C$\underline{H}$(N)-CH_2), 5.03 (broad s, 2H, -N$\underline{H}_2$), 7.81 (s, 1H, N-C$\underline{H}$=N) ppm. ^{13}C-NMR δ = 24.0, 32.6, 56.1, 125.8, 140.8, 151.3, 153.9, 158.9 ppm. MS (ESI) m/z 238.28.

Alternatively, cyclopentyl bromide (200 mg, 1.34 mmol) is added to a suspension of 6-chloroguanine in dimethylacetamide, followed by NaOMe (144 mg, 2.67 mmol). The solution is stirred over night at 100°C. The solvents are evaporated *in vacuo*. The residue is adsorbed on silica (1g) and purified by flash chromatography (ethyl acetate / petrol ether 1 : 1). Yield: 140 mg (0.59 mmol, 44%).

Example 5: N^9-Cyclopentyl-O^6-(4-bromothiophen-2-yl)-guanine (CPTG, 18)

To a solution of N^9-cyclopentyl-6-chloroguanine (17) (50 mg, 0.21 mmol) in DMF (1.3 mL) 1,4-diazabicyclo[2.2.2]octane (DABCO) (71 mg, 0.63 mmol) is added. T he reaction mixture is stirred 3 h at room temperature. TLC shows complete reaction. A solution of (4-bromothiophen-2-yl)-methanol (14) (49 mg, 0.25 mmol) and 1,8-diazabicyclo[5.4.0]-undec-7-ene (DBU) (96 mg, 0.094 mL, 0.63 mmol) in DMF (0.7 mL) is then added to the reaction mixture. The solution is stirred over night at room temperature. The crude product is purified by flash chromatography (ethyl acetate / petrol ether 1 : 9 → 3 : 7). Yield 25 mg (0.063 mmol, 30%). TLC R_f = 0.23 (ethyl acetate / petrol ether 1 : 1). ^{1}H NMR ($CDCl_3$) : δ = 1.77 (m, 2H, CH_2-C$\underline{H}_2$-CH(OH) *syn*), 1.92 (m, 4H, CH_2-C$\underline{H}_2$-CH_2 *anti*, CH_2-C$\underline{H}_2$-CH(OH) *anti*), 2.21 (m, 2H, CH_2-C$\underline{H}_2$-CH_2 *syn*), 4.76 (qnt, 1H, CH_2-C$\underline{H}$(N)-

CH_2), 4.87 (broad s, 2H, -NH_2), 5.64 (s, 2H, O-CH_2-C), 7.11 (s, 1H, C=CH-C(Br)), 7.18 (s, 1H, C(Br)-CH-S), 7.66 (s, 1H, N-CH-N) ppm. ^{13}C-NMR δ = 24.0, 32.8, 55.7, 61.8, 109.2, 116.0, 124.1, 131.0, 138.0, 140.1, 154.6, 158.7, 160.3 ppm. MS (ESI) m/z 394.36.

Example 6: N^9-Benzyl-6-chloroguanine (19)

A flask is charged with 6-chloroguanine, PPh_3 and benzyl alcohol. The mixture is dried *in vacuo* for 3 h and subsequently dissolved in dry THF to which activated molecular sieves (4 A) are added. After stirring for 15 min diisopropyl azodicarboxylate (DIAD) is added. The reaction is stirred over night. Solvent is removed under reduced pressure, and the product is purified *via* column chromatography (petrol ether / ethyl acetate 3 : 1). ^{1}H-NMR: 7.65-7.26 (m, 6H), 5.25 (s, 2H), 5.10 (s, br, 2H); UV: λ_{max}: 306 nm.

Example 7: N^9-Benzyl-O^6-(4-bromothiophen-2-yl)-guanine (20)

To a solution of N^9-benzyl-6-chloroguanine (19) (50 mg, 0.19 mmol) in DMF (1.5 mL) 1,4-diazabicyclo[2.2.2]octane (DABCO) (65 mg, 0.57 mmol) is added. The reaction mixture is stirred 3 h at room temperature. TLC shows complete reaction. A solution of (4-bromo-thiophen-2-yl)-methanol (14) (44 mg, 0.23 mmol) and 1,8-diazabicyclo[5.4.0]-undec-7-ene (DBU) (88 mg, 0.086 mL, 0.57 mmol) in DMF (0.7 mL) is then added to the reaction mixture. The solution is stirred over night at room temperature. The crude product is purified by flash chromatography (ethyl acetate / petrol ether 1 : 4 $\rightarrow$ 3 : 7) Yield 27 mg (0.065 mmol, 34%). TLC R_f = 0.20 (ethyl acetate / petrol ether 1 : 1). ^{1}H NMR ($CDCl_3$) : δ = 4.88 (broad s, 2H, -NH_2), 5.23 (s, 2H, N-CH_2-C), 5.65 (s, 2H, O-CH_2-C), 7.12 (s, 1H, C=CH-C(Br)), 7.19 (s, 1H, C(Br)-CH-S), 7.24-7.71 (m, 5H, arom.), 7.56 (s, 1H, N-CH-N) ppm. MS (ESI) m/z 416.33.

Example 8: Reaction rate of AGT mutants with N^9-cyclopentyl-O^6-(4-bromothiophen-2-yl)-guanine (18, CPTG)

A first AGT mutant Gly131Lys, Gly132Thr, Met134Leu, Arg135Ser, Asn157Gly, Ser159Glu is prepared by directed evolution. Two partially overlapping regions of the PGEG-hAGT gene, an AGT containing the mutations Asn157Gly, Ser159Glu (Juillerat *et al.*, Chem Biol 10:313-317, 2003), are amplified with the suitable primers in separate reactions. The primers contain the nucleotide mixtures NNK (N=A, C, G or T; K= G or T) at positions corresponding to the codons 131, 132, 134, 135 of the hAGT gene. With

respect to their partial complementarity, these two PCR fragments are assembled in a further PCR reaction, and amplified to give rise to full length genes randomised at codons 131, 132, 134, 135. These are cloned in fusion to the g3 protein of filamentous phage in the vector pAK100 via SfiI restriction sites. The resulting gene library is used for phage display.

Production of phages of this library is carried out in *E.coli* JM101 cells. An exponential culture is superinfected with helper phage and grown overnight at 24°C. The supernatant of this culture is incubated with 1 µM digoxigeninylated O^6-benzylguanine (substance 2 of Juillerat *et al.*, Chem Biol 10:313-317, 2003) for 6 minutes. In subsequent selection rounds, the reaction time is decreased to 90 seconds and 45 seconds, respectively, and the concentration of substrate is decreased to 10 nM to increase selection pressure. Phages are purified from this reaction by precipitation with 4% PEG / 3% NaCl. The phages carrying mutant AGT that is now covalently labeled with digoxigenin are isolated by incubation with magnetic beads coated with anti-digoxigenin antibodies (Roche Diagnostics), and used for re-infection of bacteria.

Selected AGT mutants are amplified and subsequently cloned between the BamH1 and EcoR1 sites of the expression vector pGEX-2T (Amersham). This allows the expression of the inserted gene as a C-terminal fusion to the GST protein, the gene of which is provided by the vector.

Protein expression from this vector is carried out in *E. coli* strain BL21. An exponentially growing culture is induced with 0.5 mM IPTG, and the expression is carried out for 3.5 h at 24°C.

Purification: The harvested cells are resuspended in a buffer containing 50 mM phosphate, 0.5 M NaCl, 1 mM DTT, supplemented with 1 mM PMSF and 2 µg/mL aprotinin, and disrupted by lysozyme and sonification. The cell debris are separated by centrifugation at 40000 x g. The extract is applied to pre-equilibrated glutathione sepharose (Amersham) which is then washed with 20 bed volumes (50 mM phosphate, 0.5 M NaCl, 1 mM DTT). The mutated GST-AGT fusion protein is eluted with 10 mM reduced glutathione in 50 mM Tris·HCl pH 7.9. The purified protein is dialyzed against 50 mM HEPES pH 7.2; 1 mM DTT; 30 % glycerol and then stored at –80°C.

A second AGT mutant ("AGTM") containing the mutations Cys62Ala, Gln115Ser, Gln116His, Lys125Ala, Ala127Thr, Arg128Ala, Gly131Lys, Gly132Thr, Met134Leu, Arg135Ser, Cys150Asn, Ser151Ile, Ser152Asn, Asn157Gly, Ser159Glu, and truncation at 182 is derived from the mentioned above AGT mutant by subsequent introduction of further mutations and a truncation after codon 182 into the mutant AGT gene. Subsequent PCR amplifications gives rise to the further mutated gene that is subcloned into pGEX2T as described for the first mutant. The GST-fusion protein is expressed and purified as well.

AGTM (1µM final concentration) is incubated with N^9-cyclopentyl-O^6-(4-bromothiophen-2-yl)-guanine (18, CPTG) (10 µM in 1% DMSO final concentration) in reaction buffer (50 mM HEPES, 1 mM DTT, 200 mg/mL BSA, pH 7.3) at 24°C. Samples are taken at defined times (5, 10, 15, 20, 30, 45, 60, 90, 120 min) and directly incubated with biotinylated O^6-benzylguanine (BGBT, substance 3a of Juillerat *et al.*, Chem Biol 10:313-317, 2003, 10 µM final concentration) for 4 min. The reaction is quenched by addition of SDS-Laemmli buffer and heating for 2 min at 95°C. Samples are run on a 12% acrylamide gel and analysed by Western blotting. The result is shown in Figure 2.

The second order rate constant is: $k_2 = 92.5$ [s^{-1} M^{-1}]. This value is very slow compared to the activity of AGTM with BGBT ($k_2 = \sim 3{,}000$ [s^{-1} M^{-1}]). The activity ratio of the two substrates BGBT / CPTG is at least 25.

Example 9: *In vitro* competitions between N^9-cyclopentyl-O^6-(4-bromothiophen-2-yl)-guanine (18, CPTG) and biotinylated O^6-benzylguanine (BGBT)

Mutant AGT (AGTM) and wild type human AGT (0.5 µM final concentration) are separately incubated with BGBT (0.5 µM final concentration) and different concentrations of CPTG (0, 0.5, 1, 5, 10 µM final concentrations) in reaction buffer (50 mM HEPES, 1 mM DTT, 200 mg/mL BSA, pH 7.3) for 45 min. The reaction is quenched by addition of 2x SDS buffer and placed 2 min at 95°C. Samples are run on a 12% acrylamide gel and analysed by Western blotting.

AGTM and wild type hAGT show significantly different reactivity toward CPTG in the presence of both compounds CPTG and BGBT. With a concentration of CPTG 20 times

higher than BGBT, 95% of wild type AGT is quenched by CPTG after 45 minutes while 87% of the mutant AGTM is still active. As it is known that the rate constant of the reaction of wild type hAGT with BGBT is 400 $s^{-1}M^{-1}$, it can be assumed that the rate constant of wild type hAGT with CPTG is not significantly affected based on the known reactivity of hAGT with the sterically comparable N^9-desoxyribosyl-O^6-benzylguanine which gives a value of 340 $s^{-1}M^{-1}$ (Lodewijk *et al.,* European Patent 704 445).

A solution of both AGTM-6xHis fusion protein (0.2 µM final concentration) and wild type hAGT-GST fusion protein (1.2 µM final concentration) in reaction buffer (50 mM HEPES, 1 mM DTT, 200 mg/mL BSA, pH 7.3) is incubated with BGBT (5 µM final concentration) and different concentrations of CPTG (0, 5, 10 µM final concentrations) for 30 min. The reaction is quenched by addition of 2x SDS buffer and placed 2 min at 95°C. Samples are run on a 12% acrylamide gel and analysed by Western blotting. The result is shown in Figure 3.

In comparing the reaction of AGTM and wild type hAGT, a high specificity of CPTG toward hAGT is observed. At the same concentration (both substrates 5 µM final concentration), 95% of wild type AGT has reacted with CPTG versus only 5% of AGTM.

Example 10: Fluorescence-appearance assay

Two aliquots of 100 µL hAGT-GST fusion protein in reaction buffer (50 mM Tris, 1 mM DTT, 1 mg/mL BSA, pH 8.0, final concentration of hAGT-GST: 1 µM) are placed in a black flat-bottomed microtiter plate. One aliquot serves as a control, which is quenched with benzylguanine (final concentration of 100 µM) for 40 min. Compound (7) in DMSO is added to both samples of hAGT-GST (final concentration 10 nM) and incubated at 25°C for 1 hour. The fluorescence is read on a Victor^2 V plate reader (Perkin-Elmer) with the fluorescein-protocol. The fluorescence reading in the sample is at least twofold higher than in the control.

Example 11: One-pot labeling and separation assay

50 µL hAGT-GFP fusion protein in reaction buffer (50 mM Tris, 1 mM DTT, 1 mg/mL BSA, pH 7.3, final concentration of hAGT-GFP: 5 µM) is incubated with substrate (9) (B=biotin, final concentration 10 µM) for 1 h. 150 µL streptavidin coated beads (Dynabeads M-280) are added and the incubation time and magnetic separation from

the supernatant are done according to manufacturer's instruction for use. By measuring the GFP absorption at 488 nm, the amount of fusion protein in solution before and after the treatment with magnetic beads is determined. The loss of fusion protein is less than 30%. The labeling efficiency is determined by measuring the absorption of the ALEXA-dye at 594 nm in the supernatant after removal of the unreacted substrate. The labeling efficiency is at least 20%.

Example 12: Labeling beads

For the preparation of substrate related to (9) carrying carboxyfluoresceine as the label L_1, NHS-activated Sepharose™ packed into a polypropylene column (HiTrap™, 1 mL column volume, Amersham Biosciences) is incubated with a solution of N^9-(3-amino-propyl)-4-(5(6)-carboxyfluorescein-aminomethyl)benzylguanine (corresponds to compound (6), but carrying a different N^9 substituent, 8 mM) in 0.5 M NaCl, pH 8.5, for 30 min. The column is washed first with 6 ml of Buffer A (0.5 M ethanolamine, 0.5 M NaCl, pH 8.3), followed by 6 ml of Buffer B (0.1 M acetic acid, 0.5 M NaCl, pH 4). To quench any remaining NHS esters the column is subsequently incubated with Buffer A for 30 min and then washed again with 6 ml of Buffer A, followed by 6 ml of Buffer B and then again 6 ml of Buffer A. 100 µL functionalized Sepharose beads are equilibrated by washing with 5 ml of Buffer C (10 mM HEPES, pH 7.4, 150 mM NaCl, 0.05% Tween 20). To this suspension are added 100 µL of hAGT-GFP (final concentration of 50 µM) and incubated at 25°C for 1 h. The sepharose beads are removed by centrifugation and the labeling efficiency is determined by measuring the absorbtion of the fluorescein at 488 nm in the supernatant. The labeling efficiency is at least 20%. By measuring the GFP absorption at 488 nm, the amount of fusion protein in solution before and after the treatment with modified sepharose beads is determined. The loss of fusion protein is less than 30%.

Claims

1. A compound of formula (1)

$$R_1–A–X–CH_2–R_3–R_4–L_1 \qquad (1)$$

wherein
A is a group recognized by O^6-alkylguanine-DNA alkyltransferases (AGT) as a substrate;
X is oxygen or sulfur;
R_1 is a group $–R_2–L_2$ or a group R_5;
R_2 and R_4 are, independently of each other, a linker;
R_3 is an aromatic or a heteroaromatic group, or an optionally substituted unsaturated alkyl, cycloalkyl or heterocyclyl group with the double bond connected to CH_2;
R_5 is arylmethyl or heteroarylmethyl or an optionally substituted cycloalkyl, cycloalkenyl or heterocyclyl group;
L_1 is a label, a plurality of same or different labels, a bond connecting R_4 to A forming a cyclic substrate, or a further group $–R_3–CH_2–X–A–R_1$; and
L_2 is a label or a plurality of same or different labels.

2. The compound according to claim 1 of formula (1) wherein
A is a heteroaromatic group containing 1 to 5 nitrogen atoms;
X is oxygen;
R_1 is a group $–R_2–L_2$ or a group R_5;

R_2 and R_4 are, independently of each other, a straight or branched chain alkylene group with 1 to 300 carbon atoms, wherein optionally
(a) one or more carbon atoms are replaced by oxygen, in particular wherein every third carbon atom is replaced by oxygen, e.g. a poylethyleneoxy group with 1 to 100 ethyleneoxy units;
(b) one or more carbon atoms are replaced by nitrogen carrying a hydrogen atom, and the adjacent carbon atoms are substituted by oxo, representing an amide function –NH–CO–;
(c) one or more carbon atoms are replaced by oxygen, and the adjacent carbon atoms are substituted by oxo, representing an ester function –O–CO–;

(d) the bond between two adjacent carbon atoms is a double or a triple bond, representing a function –CH=CH– or –C≡C–;
(e) one or more carbon atoms are replaced by a phenylene, a saturated or unsaturated cycloalkylene, a saturated or unsaturated bicycloalkylene, a bridging heteroaromatic or a bridging saturated or unsaturated heterocyclyl group;
(f) two adjacent carbon atoms are replaced by a disulfide linkage –S–S–;
or a combination of two or more, especially two or three, alkylene and/or modified alkylene groups as defined under (a) to (f) hereinbefore, optionally containing substituents;

R_3 is phenyl, an unsubstituted or substituted mono- or bicyclic heteroaryl group of 5 or 6 rings atoms comprising zero, one, two, three or four ring nitrogen atoms and zero or one oxygen atom and zero or one sulfur atom, with the proviso that at least one ring carbon atom is replaced by a nitrogen, oxygen or sulfur atom, 1-alkenyl, 1-alkinyl, 1-cyclohexenyl with 3 to 7 carbon atoms, or an optionally substituted unsaturated heterocyclyl group with 3 to 12 atoms and 1 to 5 heteroatoms selected from nitrogen, oxygen and sulfur, and a double bond in the position connecting the heterocyclyl group to methylene CH_2;

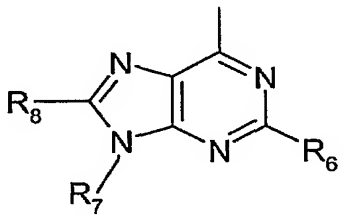
R_5 is optionally substituted phenylmethyl or naphthylmethyl; optionally substituted heteroarylmethyl wherein heteroaryl is a mono- or bicyclic heteroaryl group comprising zero, one, two, three or four ring nitrogen atoms and zero or one oxygen atom and zero or one sulfur atom, with the proviso that at least one ring carbon atom is replaced by a nitrogen, oxygen or sulfur atom, and which has 5 to 12 ring atoms; optionally substituted cycloalkyl with 3 to 7 carbon atoms; optionally substituted cycloalkenyl with 5 to 7 carbon atoms; optionally substituted saturated or unsaturated heterocyclyl with 3 to 12 atoms, and 1 to 5 heteroatoms selected from nitrogen, oxygen and sulfur;

L_1 is one or a plurality of same or different labels selected from a spectroscopic probe, a magnetic probe, a contrast reagent, a molecule which is one part of a specific binding pair which is capable of specifically binding to a partner, a molecule that is suspected to interact with other biomolecules, a library of molecules that are suspected to interact with other biomolecules, a molecule which is capable of crosslinking to other molecules, a molecule which is capable of generating hydroxyl radicals upon exposure to H_2O_2 and

ascorbate, a molecule which is capable of generating reactive radicals upon irradiation with light, a molecule covalently attached to a solid support, a nucleic acid or a derivative thereof capable of undergoing base-pairing with its complementary strand, a lipid or other hydrophobic molecule with membrane-inserting properties, a biomolecule with desirable enzymatic, chemical or physical properties, a bond connecting R_4 to A forming a cyclic substrate, and a further group $-R_3-CH_2-X-A-R_1$; and

L_2 is one or a plurality of same or different labels selected from a spectroscopic probe, a magnetic probe, a contrast reagent, a molecule which is one part of a specific binding pair which is capable of specifically binding to a partner, a molecule that is suspected to interact with other biomolecules, a library of molecules that are suspected to interact with other biomolecules, a molecule which is capable of crosslinking to other molecules, a molecule which is capable of generating hydroxyl radicals upon exposure to H_2O_2 and ascorbate, a molecule which is capable of generating reactive radicals upon irradiation with light, a molecule covalently attached to a solid support, a lipid or other hydrophobic molecule with membrane-inserting properties, and a biomolecule with desirable enzymatic, chemical or physical properties.

3. The compound according to claim 1 of formula (1) wherein the group R_1-A is a purine radical of formula (2)

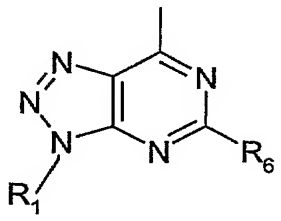


(2)

wherein R_6 is hydrogen, hydroxy or unsubstituted or substituted amino; and one of R_7 and R_8 is R_1 and the other one is hydrogen.

4. The compound according to claim 3 of formula (1) wherein the group R_1-A is a purine radical of formula (2), R_6 is unsubstituted amino, R_7 is R_1, R_8 is hydrogen, and X is oxygen.

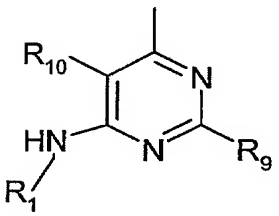
5. The compound according to claim 1 of formula (1) wherein the group R_1–A is an 8-azapurine radical of formula (3)



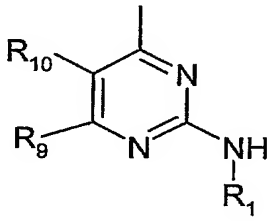
(3)

wherein the substituents R_6 is hydrogen, hydroxy or unsubstituted or substituted amino.

6. The compound according to claim 1 of formula (1) wherein the group R_1–A is a pyrimidine radical of formula (4a) or (4b)



(4a)



(4b)

wherein R_9 is hydrogen, halogen, lower alkyl with 1 to 4 carbon atom or amino, and R_{10} is hydrogen, halogen, lower alkyl with 1 to 4 carbon atoms, amino, nitro or nitroso.

7. A method for detecting and/or manipulating a protein of interest, wherein the protein of interest is incorporated into an AGT fusion protein, the AGT fusion protein is contacted with a compound of formula (1) according to any one of claims 1 to 6, and the AGT fusion protein is detected and optionally further manipulated using the label L_1 in a system designed for recognising and/or handling the label.

8. The method according to claim 7, wherein in the compound of formula (1) label L_2 is a solid support, and the AGT fusion protein contacted with the compound of formula (1) is separated from the compound of formula (1) by filtration or centrifugation or separation of magnetic beads.

9. The method according to claim 7, wherein in the compound of formula (1) label L_1 is one member and label L_2 the other member of two interacting spectroscopic probes L_1 / L_2, and the AGT fusion protein is detected by fluorescence.

10. The method according to claim 7 for detecting and/or manipulating a protein of interest, wherein the protein of interest is fused with a mutant AGT, the mutant AGT fusion protein is contacted with a mixture of
(a) a compound of formula (1) wherein R_1 is a group R_5 and which is not recognized by the mutant AGT, and
(b) another compound of formula (1) recognized by the mutant AGT fusion protein,
and the mutant AGT fusion protein is detected and optionally further manipulated using the label in a system designed for recognizing and/or handling the label.

Abstract

The invention relates to substrates for O^6-alkylguanine-DNA alkyltransferases (AGT) of formula R_1–A–X–CH_2–R_3–R_4–L_1, wherein A is a group recognized by AGT as a substrate, X is oxygen or sulfur, R_1 is a group –R_2–L_2 or a group R_5, R_2 and R_4 are, independently of each other, a linker, R_3 is an aromatic or a heteroaromatic group, or an optionally substituted unsaturated alkyl, cycloalkyl or heterocyclyl group with the double bond connected to CH_2, R_5 is arylmethyl or heteroarylmethyl or an optionally substituted cycloalkyl, cycloalkenyl or heterocyclyl group, L_1 is a label, a plurality of same or different labels, a bond connecting R_4 to A forming a cyclic substrate, or a further group –R_3–CH_2–X–A–R_1, and L_2 is a label or a plurality of same or different labels. The invention further relates to methods of transferring a label from these substrates to O^6-alkylguanine-DNA alkyltransferases (AGT) and AGT fusion proteins.

Fig. 1

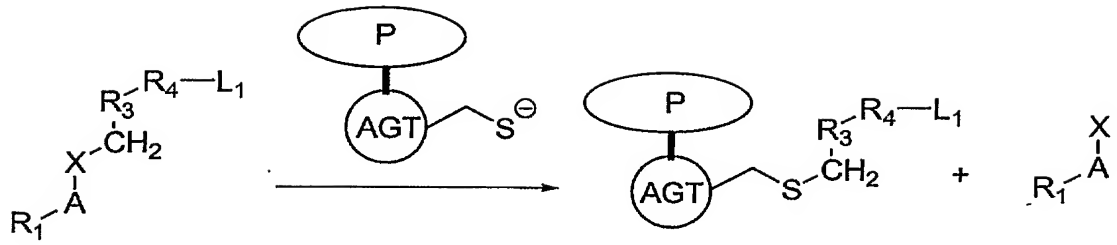


Fig. 2:

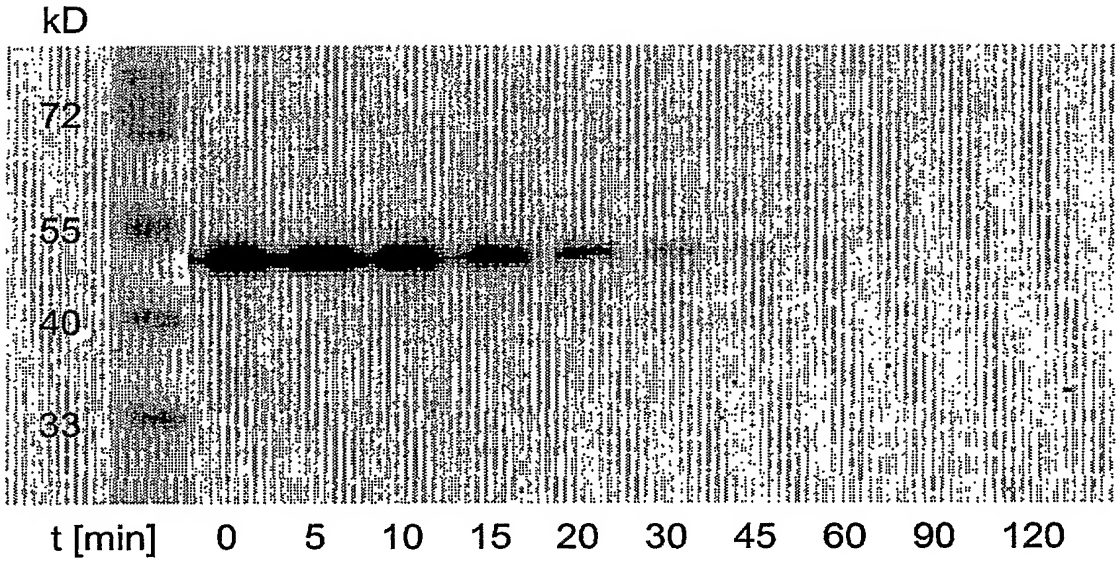


Fig. 3: